

Another one bites the dust

The hazards of seeking to implement reforms at universities with outstanding reputations have been demonstrated once again, this time in Switzerland.

Switzerland's flagship university, the Swiss Federal Institute of Technology (ETH) in Zurich, is arguably the strongest in mainland Europe. The government already provides one of the highest per-capita expenditures on research in the world, and is set to increase funding by a further 30% over the next five years.

But if nothing seems to be broken, does anything need fixing? Ernst Hafen, the molecular biologist who became president of the ETH last December, thought it did. In his view, major reforms were required in order to safeguard the university's pre-eminence. Unlike most Swiss universities, where deans share decision-making, the ETH president's position is a powerful one and, in theory, Hafen should have been able to have his way.

But the ETH faculty put paid to that, pulling rank after a short but bitter public dispute and securing Hafen's resignation just eleven months in (see page 130). The hostility of the ETH faculty to the man who wanted to restructure it echoes that attracted by Larry Summers at Harvard University two years ago (see *Nature* **433**, 190–192; 2005), which culminated in a row over his unfortunate comments on the aptitude of women and his subsequent resignation.

Hafen had been impressed by a previous, successful reorganization five years ago at the other federally funded university in Switzerland, the EPFL Lausanne — the remaining universities in Switzerland are funded by their local cantons — and wanted to introduce a more corporate management model at the ETH.

The ETH currently has little hierarchy below the president, and what there is mostly concentrates in the *Schulleitung*, comprising the president, two vice-presidents (responsible for research and infrastructure) and the faculty-elected rector (responsible for teaching). The committee of department heads meets quarterly with the president to exchange information.

Hafen wanted to merge some departments into a less unwieldy

number of schools led by professional deans. More contentiously, he sought to reform the *Schulleitung*, eliminating the rector's position and introducing five vice-presidents, whose arrival might have threatened the clout of senior academics.

University professors are notoriously conservative and jealous of their local powers — especially at elite institutions such as the ETH. There might therefore be a tendency to dismiss the brouhaha as academic provincialism. But this would be wrong for two reasons.

First, the ETH professors have in the recent past already cooperated with changes implemented by the president's office. That was a major reason for their reluctance this time: they are still in the process of enacting the last round of reforms, which made departments responsible for their own finances and put in place a new teaching system. They were clearly unconvinced that the time was ripe for reform of these reforms.

And therein lies the second reason. Hafen failed to convince the professors that his proposals made sense, and he declined to engage in the dialogue that might have won them over.

Perhaps too confident of the power conferred on him on paper, he failed to talk and, fatally, he failed to listen. He pushed too fast initially and, when an impasse arose, he moved swiftly into reverse, withdrawing his whole reform package in its entirety on 23 October, thereby losing further respect.

The ETH will doubtless survive this embarrassing setback. But Hafen's departure leaves questions about whether his proposals were really being driven by the ETH supervisory council, which hired him but failed to back him when the chips were down. The council must now start afresh, by appointing a president who will inspire the confidence of the university staff. ■

"Hafen wanted to introduce a more corporate management model at the ETH."

Correction or retraction?

Errors reported in this issue by authors of a *Nature* paper pose a dilemma about trust.

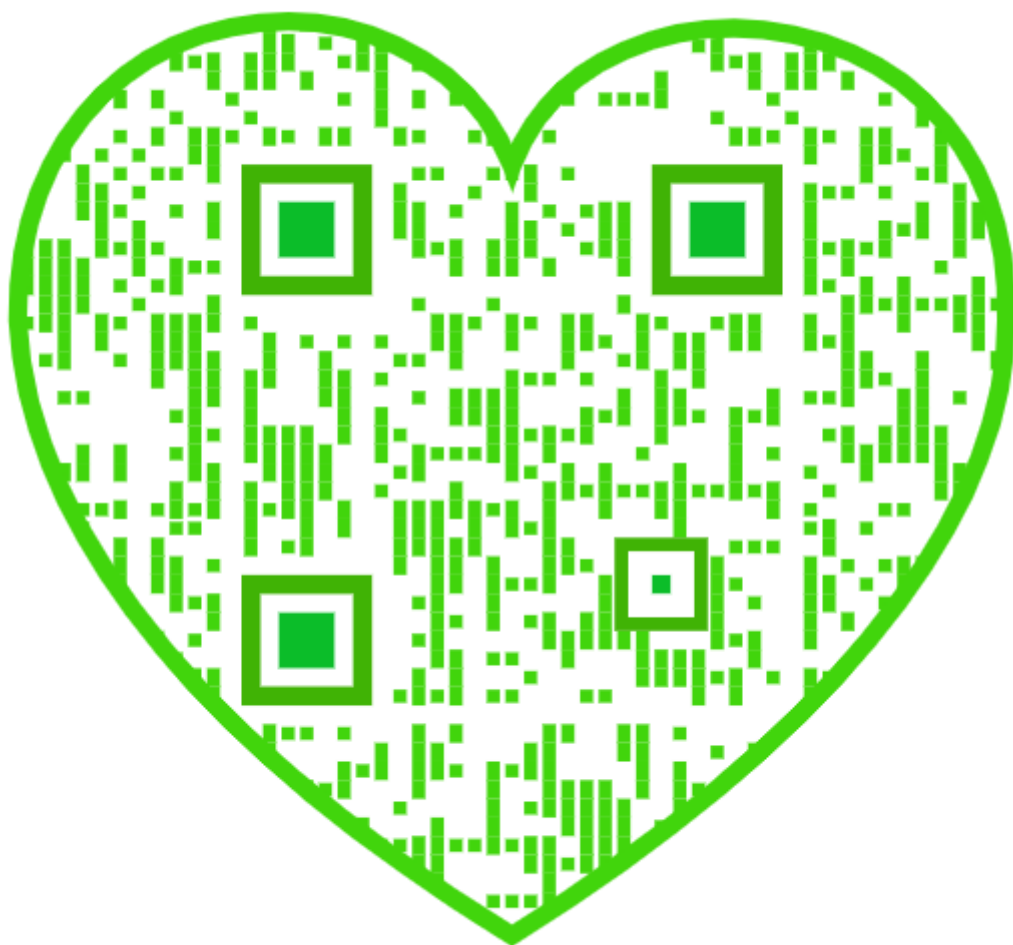
The practice of science — and the publication of science in particular — is often lauded for its capacity for self-correction, and, to a large extent, deservedly so. During the peer-review process, basic mistakes and errors of judgement are frequently identified, minimizing the number that make it through into often much-improved final publications. Erroneous results that do slip through the peer-review net may be promptly identified; if sufficiently serious, they are corrected or even withdrawn.

The reality of science publication does not always accord with such

idealistic expectations, however, and the correction published this week (on page 235) is a case in point.

When an important error has been made in a published piece of work, yet the central claim or result still stands, the publication of a Corrigendum is the most sensible way forward; others are thereby alerted to any inaccuracies in the paper, which may have an impact on their own research. If, on the other hand, the errors that have been made undermine the principal message of the paper, then a retraction is in order — the paper may still contain valid scientific information, but the original publication has now lost its *raison d'être*.

There is a grey area in between, exemplified by the events — past and present — that have now culminated in the aforementioned Corrigendum. The original paper (*Nature* **366**, 143–146; 1993) is viewed by many as a landmark in its field: an experimental 'first', in which compositional analysis at atomic resolution had been achieved



with an electron microscope. But the paper did not have an easy time with referees, with one reviewer in particular maintaining that there were “disquieting questions” over both the provenance of key data and inconsistencies in the manner in which they had been subsequently reprocessed (in response to concerns raised earlier during the review process) in order to better substantiate the central claim. The authors, however, had provided *Nature* with a robust defence of their work in response to these remaining criticisms — firm assurances that at the time we accepted at face value. Combined with a positive endorsement of the work that was offered by a second referee, we felt it appropriate to proceed with publication of the paper.

Seen with the benefit of hindsight, we made an error of judgement in taking those assurances on trust — perhaps influenced by our awareness of intense competition in the field at the time. For it has belatedly transpired that many of the more critical referee’s “disquieting questions” did indeed have a sound technical basis, as key data were misrepresented by the authors, both during the review process and in the final published version of the paper. This has now been fully acknowledged by the authors, who say that they are mystified as to why they offered the assurances they did when originally challenged on these points.

So what does this mean for the paper? It could be argued, as we have done during extensive deliberations, that confidence in the validity of the work has been severely compromised as a consequence of these errors, and that the paper should be formally retracted. But we have concluded for several reasons that the authors’ Corrigendum should

be accepted. Thirteen years have elapsed since publication, and the authors no longer have available the source data to reanalyse and resubmit to peer review — and so are not in a position to mount a thorough defence of their published results against the remaining criticisms. Furthermore, an investigation commissioned by Oak Ridge National Laboratory in Tennessee has robustly cleared the authors of any intent to deceive; we have received assurances that the original data, if consistently analysed as intended, would still have supported the central thesis of the paper. And most importantly, the authors have both acknowledged and proposed corrections for their earlier mistakes.

In the end, it comes down to an issue that is at the very heart of the practice and communication of science: the question of trust. After all, if researchers and editors cannot safely assume, even as a starting point, that scientific results are essentially true as reported, then the advancement of science is in serious trouble.

Without doubt, there has been in this case a severe breach of the trust on which the publication of science is based. But the reasons underlying it, and the hypothetical outcome for the work had these concerns been tackled more robustly when first raised, can now only be speculated about. Other researchers will have their own take on the situation (see page 129).

There is some consolation in the fact that the experimental capability first reported in the flawed *Nature* paper was soon exceeded, as reported by these authors and others, in a range of different contexts. Thus we can at least be relieved that the progress of science was not impeded by this particular episode. ■

Smart but lightweight

An imaginative innovation policy in Britain continues to be under-resourced.

It is the intention of every government on Earth to inspire increased research, development and innovation in the private sector. How this should actually be done remains something of a mystery, however. Last week, the British government took a stab at the problem, announcing an administrative change that it hopes will help it meet its ambitious stated goal of expanding business expenditure on research and development from 1.2% of the economy to 1.7% by 2014.

Innovation policy in Britain traditionally falls under the remit of the Department of Trade and Industry (DTI), which has long housed a pot-pourri of small initiatives and programmes aimed at fostering industrial innovation. Two years ago, this mixture was placed under the guidance of the Technology Strategy Board, an advisory committee chaired by Graham Spittle of IBM.

Early next year, the Technology Strategy Board will effectively be spun off from the DTI and reconstituted as an autonomous entity. It will hire a chief executive and a small staff, and will operate at arm’s length from the government, managed by a board drawing members from industry and finance. The group will be constituted much like one of the research councils that support British scientific research, but will have a different mission — fostering innovation in business.

The change will give the group more latitude for effective action, and pulling it out of the DTI will have at least two other significant advantages. It will help the committee to address innovation, not just in manufacturing industry (the DTI’s traditional remit) but also in the service sector — ranging from publishing to banking — which now constitutes four-fifths of Britain’s economy. And the free-standing committee will be better placed to work with all departments of government, and address the area where the state can arguably make the greatest difference of all, by supporting innovative suppliers through the £150 billion (US\$290 billion) or so that it spends each year on goods and services.

However, the resources under the direct control of the new body will remain, in the first instance, rather paltry. Around £170 million worth of grants and other programmes is, for all the talk about intelligent leverage, unlikely to spur much of anything across an economy the size of Britain’s.

The Confederation of British Industry — which has, of course, an interest in the matter — has advocated the spin-off of the Technology Strategy Board, but suggested that it needs four times as much money to have an impact. The question of additional resources will be addressed by the Treasury in next year’s Comprehensive Spending Review. Its outcome will tell us how much faith the government has in this particular innovation. ■

“The Technology Strategy Board will be constituted much like a research council, but will have a different mission — fostering innovation in business.”

RESEARCH HIGHLIGHTS

Amazon given a good dust

Environ. Res. Lett. doi:10.1088/1748-9326/1/1/014005 (2006)

Millions of tonnes of dust from the Sahara blow across the Atlantic every year and fertilize the Amazon rainforest. Now the origin of much of that dust has been identified as a small region of northern Chad.

Ilan Koren of the Weizmann Institute in Rehovot, Israel, and his co-workers used satellite data to track the size and mass of dust plumes leaving the Sahara. They estimate that 56% of the 40 million tonnes that is deposited in the Amazon basin annually originates from the Bodélé depression, a region with an area of less than 45,000 square kilometres.

The Bodélé is a productive source because the region's fine diatomite soils are frequently eroded by winds channelled between a gap in two nearby mountain ranges.



J. GILES

IMMUNOLOGY

Hot-blooded

Nature Immunol. doi:10.1038/ni1406 (2006)

The next time you feel feverish, think twice before reaching for the aspirin. The febrile response has been conserved over 450 million years of evolution, and a new study reveals some of the physiological benefits that may explain why it persists.

Sharon Evans of the Roswell Park Cancer Institute in Buffalo, New York, and her colleagues show in mice that feverish temperatures may help white blood cells known as lymphocytes to reach lymph nodes. Here they are more likely to encounter, and deal with, troublesome pathogens. The team found that 'gatekeeper' cells that line blood vessels have more adhesion molecules on their surface when their temperature is raised. This would help the gatekeepers to grab blood-borne lymphocytes and pull them into the nodes.

ELECTROPHYSIOLOGY

Unusual suspects

J. Neurosci. 26, 10992-11000 (2006)

The identity of the protein machine that detects sound in our ears has long been elusive. Now a team of researchers has composed the molecular equivalent of a photofit to narrow down the search.

Cells in the inner part of the mammalian ear detect sound when vibrations bend their tiny hair-like cilia. An ion channel (as yet unidentified at the molecular level) sits in the cilia's membranes and opens when the bending tugs on fine tethers called tip links roped between the cilia.

Robert Fettiplace of the University of Wisconsin Medical School in Madison and his colleagues studied the electrical conductance properties of individual channels in rats. They found that there are two channels per tip link, and suggest that the channels probably exist in several forms with different conductances. Few known channels on the 'suspect list' have similar properties.

ARCHAEOLOGY

Skull and crossed bones

Proc. Natl Acad. Sci. USA doi:10.1073/pnas.0608443103 (2006)

Ancient bones that show a mixture of modern human and Neanderthal features support the idea that the two groups interbred, argue Erik Trinkaus of Washington University in St Louis, Missouri, and his colleagues.

Trinkaus's team reanalysed 30,000-year-old remains that were unearthed in a Romanian cave more than 50 years ago. The remains include a skull (pictured below) that has the small brow bone of a modern human and a pronounced bump on the back that is more typical of Neanderthals.



Genetic studies have so far revealed no evidence of interbreeding between the groups. But Trinkaus points to these remains and two other sets that he interprets as hybrids, as proof that humans and Neanderthals did more than just live side-by-side.

CELL BIOLOGY

Fat and frizzled

Development 133, 4561-4572 (2006)

Three scientists are challenging the prevailing molecular explanation for how cells in flat layers such as the skin arrange themselves.

Current theories hold that two sets of genes encoding different proteins act sequentially in a single pathway to line up, or polarize, the cells in planar epithelia.

But Peter Lawrence and José Casal of the Laboratory of Molecular Biology in Cambridge, UK, and Gary Struhl from the University of Columbia in New York show in the fruitfly *Drosophila* that the two sets of genes, exotically known as Dachsoos/Fat and Starry night/Frizzled, operate independently of each other. They also show that either set alone is sufficient to ensure planar polarization.

BIOPHYSICS

Kinky genes

Nature Nanotechnol. 1, 137-141 (2006)

DNA is bendier than previously thought.

So say Philip Nelson of the University of Pennsylvania, Philadelphia, and his colleagues, who used an atomic-force microscope to look at a DNA double helix. They observed more sharp bends than are

PNAS

PNAS

predicted by the usual 'worm-like chain' model of DNA elasticity, which assumes that the molecule deforms like a segmented elastic rod, obeying Hooke's law.

The discrepancy only shows up over distances of a few nanometres or so, but this is precisely the scale that is biologically relevant. One consequence is that, if a protein bends DNA when it binds, this should require less energy than the old picture would have implied.

ASTRONOMY

Radio collision report

Science **314**, 791–794 (2006)

Massive galactic clusters are thought to grow by accumulating smaller groups of galaxies, just as businesses grow through mergers. Now Joydeep Bagchi of the Inter-University Centre for Astronomy and Astrophysics in Pune, India, and his colleagues may have observed the after-effects of such a cosmic takeover.

The astronomers spotted two great arcs of radio emission around the galactic cluster Abell 3376. They suggest that these trace a shock front, where hot gas expelled from the cluster in a collision a billion years ago meets the colder intergalactic medium. X-ray observations show a bullet-shaped region of hot gas — remnants of the colliding galaxies — piercing the centre of the cluster.

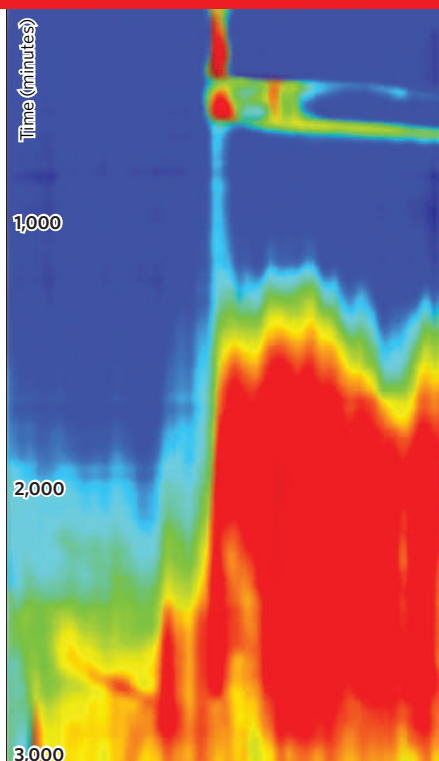
STEM CELLS

Library self-renewals

Proc. Natl Acad. Sci. USA doi:10.1073/pnas.0608156103 (2006)

Chemical libraries can be an invaluable resource for stem-cell research, reports a team led by Peter Schultz and Sheng Ding at The Scripps Research Institute in La Jolla, California.

The researchers measured the effects



of different compounds from a library of 50,000 small molecules on mouse embryonic stem cells. They found one chemical, called SC1, that maintains stem cells in an undifferentiated state. These stem cells retained their full potential — they could later be differentiated into cells from all three primary layers of the developing embryo.

The team found that SC1 interacts with the protein kinase ERK1 and an enzyme-activating protein known as RasGAP. Further investigations of SC1's role may lend insight into the mechanisms of stem-cell self-renewal.

NUCLEAR PHYSICS

Stars in the lab

Phys. Rev. Lett. **97**, 173401 (2006)

Two theorists at Tel Aviv University in Israel calculate that a tabletop fusion scheme could reproduce the nuclear reactions that take place inside stars.

Isidore Last and Joshua Jortner propose

using nanodroplets of methane (CH_4), water (H_2O) and ammonia (NH_3) as fuel for their nuclear-fusion device. Intense, femtosecond laser pulses would be used to completely ionize the molecular fuel. The hydrogen, nitrogen, carbon and oxygen nuclei should then fuse through a dynamic process known as 'Coulomb explosion'.

Laser-driven Coulomb explosion has previously been shown to trigger fusion of deuterium nuclei. If it also works for heavier nuclei, as Last and Jortner predict, the set-up would reproduce the nuclear reactions that go on inside so-called carbon–nitrogen–oxygen stars.

ECOLOGY

Evolution on a chip

Proc. Natl Acad. Sci. USA doi:10.1073/pnas.0607971103 (2006)

When evolutionary biologists talk of 'fitness landscapes', they are usually speaking metaphorically. But Robert Austin and his colleagues at Princeton University, New Jersey, have realized the concept.

The team made a row of microscopic chambers on a silicon chip in which *Escherichia coli* may grow. The bacteria can move between the chambers along narrow corridors. Adjusting the nutrient supply to each of these patches creates gradients in habitability, representing a fitness landscape that can be made arbitrarily 'smooth' or 'rugged'.

The bacteria form complex patterns as they colonize or abandon patches, adapting either genetically or physiologically to the different environments. The picture above shows how the density of bacteria in a row of 85 patches, increasing from blue through green to red, evolves over time. This ecosystem-on-a-chip should provide a useful tool for studying evolutionary dynamics.

JOURNAL CLUB

Frances Ashcroft
University of Oxford, UK

A physiologist discusses matters close to the heart.

This time last year my father was suffering from congestive heart failure. He became increasingly frail, slowing down like an unwound clockspring until, in February, his heart simply stopped.

As a physiologist, I had some idea of his condition, but I did not then realize how close it was to

my own research area.

In 1983, ATP-sensitive potassium (K_{ATP}) channels were found in the heart. These channels are gated pores that control potassium fluxes across the cell membrane. However, their precise role in the heart was unclear.

One year later, I discovered that these channels are central to the mechanism by which glucose stimulates insulin secretion from the pancreas. Unravelling the role of K_{ATP} channels in diabetes, and the way in which channel

structure influences function, has been an all-consuming passion for me ever since.

To my surprise, it now turns out that these channels also play a role in heart failure. Heart failure is usually caused by narrowing of the arteries, which increases the pressure against which the heart has to pump, making it work harder. Eventually, it fails.

Recently, Andre Terzic of the Mayo Clinic in Rochester, Minnesota, and his group showed that K_{ATP} channels confer protection against heart failure

(S. Yamada *et al.* *J. Physiol. Lond.* published online doi:10.1113/jphysiol.2006.119511; 2006). In normal mice, cardiac K_{ATP} channels open in response to an increased pressure load, reducing stress on the heart. Mice lacking K_{ATP} channels rapidly develop heart failure and die.

In the pancreas, K_{ATP} -channel activity is finely balanced: too much causes diabetes and too little hyperinsulinism. But in the heart, as this paper shows, opening is almost always beneficial.

NEWS



Lights out: the Arecibo radio telescope in Puerto Rico is one of several facilities that could face budget cuts to secure funding for larger projects.

Telescope is no longer dish of the day

With a history of nurturing Nobel prizewinners and landing star roles in Hollywood films, the Arecibo radio telescope in Puerto Rico has worked its way into both scientific and popular cultures. Since 1963, the telescope's 70,000-square-metre collecting area has been used for everything from characterizing pulsars to the finale of the James Bond film *Goldeneye*.

But, according to a review committee, the facility should close if it cannot find outside financial support by 2011. The committee delivered its verdict to the Division of Astronomical Sciences of the US National Science Foundation (NSF) on 3 November. Arecibo is not the only facility facing the axe — the committee also recommended that the continent-wide Very Large Baseline Array should seek external funding or face closure. In addition, funding for operations by the National Solar Observatory (NSO) on Kitt Peak near Tucson, Arizona, and Sacramento Peak in Sunspot, New Mexico, could end if the planned Advanced Technology Solar Telescope receives final approval.

None of the proposals is a death sentence; the NSF will consider the recommendations over the next couple of months. But the observatories in question may now turn to universities and foreign collaborators for funding. "Nobody's out here nailing plywood over the windows," says David Dooling, a spokesman for the NSO. "We think there are still a lot of years left."

But the committee's findings highlight a growing problem in US ground-based astronomy: the expanding appetite for new telescopes doesn't leave much room for the continuing operations budgets of older telescopes such as Arecibo. Among other new projects, the NSF will pay \$500 million for its share of the Atacama Large Millimeter Array, a 50-antenna cluster scheduled to be operational in Chile by 2013. The foundation also aims to participate in the Square Kilometre Array, an international effort likely to cost more than a billion dollars in total.

"To make room for the new you have to trim some of the old," says Michael Turner, an astrophysicist at the University of Chicago. "If you keep building new facilities without closing old facilities, eventually the operations budget eats your entire budget." To that end, the committee — led by astronomer Roger Blandford of Stanford University — was asked in October 2005 to find ways of saving \$30 million in the NSF's astronomy facilities budget by 2011.

The committee report recommends that astronomical facilities "plan for a natural life cycle ending, in every case, with closure".

Turner says that ground-based astronomy has so far been spared difficult choices over which facilities to close, unlike the decommissioning of space telescopes: "This is much

more difficult to do on the ground because space facilities fall from the sky and don't live in anyone's congressional district." (This might surprise Maryland Senator Barbara Mikulski, whose support of life-extending missions to the Hubble Space Telescope has been appreciated at the Space Telescope Science Institute in Baltimore, Maryland.)

Donald Campbell, an astronomer at Cornell University in Ithaca, New York, who has used Arecibo to study planets and the Moon, says that the telescope's potential closure is particularly painful in light of

major upgrades, the latest in 1997. If Arecibo closes, he says he will have to rely entirely on the less-sensitive Goldstone radar telescope operated by NASA in California's Mojave Desert. "It's a great shame that the choice has to come down to closing facilities that are unique," he says.

Observers overseas agree. "It's a pity that they were under pressure to make these cuts in facilities that were still productive by the standards of other scientific facilities," says Martin Rees, a cosmologist at the University of Cambridge, UK, and president of the Royal Society. Rees says that Britain faced a similar decision when it elected to join the European Southern Observatory, and funding pressures forced the country to close some smaller telescopes.

Heidi Ledford



US MID-TERM ELECTIONS

Follow the latest developments online
www.nature.com/news/specials/uselection2006

Data handling causes image problem for top lab

Two papers — one published more than a decade ago and the other a preprint from last year — have raised doubts among experts about the work of a world-renowned microscopy laboratory.

In a Corrigendum in this week's issue of *Nature* (see page 235), three authors from an atomic-imaging group at Oak Ridge National Laboratory in Tennessee admit to treating critical data inconsistently in a paper from 1993 (ref. 1). Two of the three authors are also co-authors on an August 2005 paper, on the arxiv.org preprint server², in which some data were reproduced from one of the group's earlier papers³ without citation and another figure included duplicated data points on two sides of a plot. The authors posted a revised version, with added citations and a replacement figure, on 15 May 2006 (ref. 4).

An investigation over the summer by an independent panel, formed at the behest of the Oak Ridge lab, found "no evidence of research misconduct or fraud in either paper", according to James Roberto, the lab's deputy director for science and technology.

But within the community, doubts linger. "It's obvious that some games are being played," comments John Silcox of Cornell University in Ithaca, New York, who was a reviewer of the 1993 *Nature* article. "I don't trust their work."

Few researchers were willing to go on the record, but several told *Nature* privately that they also were concerned about the papers.

The group is led by Stephen Pennycook, a physicist who has been at Oak Ridge since 1982. His group helped pioneer the scanning transmission electron microscope (STEM), which uses a narrow beam of electrons to simultaneously take spectra of and image materials. The microscopes cost millions of dollars and the field is intensely competitive. Pennycook's group is at the forefront, says John Spence, a physicist at Arizona State University in Tempe: "They have the



Questions have arisen over data from a team at Oak Ridge National Laboratory.

best machine, the most money, and very good people."

In the *Nature* paper, the collaboration reported taking spectra and images of a boundary between a thin wafer of cobalt silicide and one of silicon. At the time, imaging such an interface was difficult. The team proved it was possible by showing clear spectra of both the cobalt silicide and silicon.

But in the Corrigendum, the authors admit that the cobalt silicide spectra were reproduced from an earlier conference paper by the same group⁵. The admission contradicts the authors' response to a 1993 referee report (obtained independently and verified by *Nature*'s news team) in which they deny that the spectra came from the proceedings. Pennycook says that the original statements were made because the group was confused about the referee's criticisms.

The duplication of some but not all spectra meant that different methods were used to process the background noise on either side of the boundary. Some researchers are unfazed by the error: "Clearly it's a minor mistake, but fundamentally it has no influence on the results," says Christian Colliex, head of the STEM group at the University of Paris South, who collaborates with Pennycook but had no role in the 1993 paper. But others believe the

mistake should invalidate the result. "I don't understand why they haven't retracted this paper," says Silcox. *Nature*'s view is given on page 123.

The more recent arxiv.org paper is also a study of the boundary between two materials: one ferromagnetic and one superconducting. In the first version of the paper², several graphs and one image appear identical to those in an earlier paper by the group³. In addition, a plot in the paper contains identical but mirrored data on its right- and left-hand sides. "It looks fishy," says David Muller, a physicist at Cornell, who obtained his PhD under Silcox.

The paper was first posted at arxiv.org on 23 August 2005. Others in the field became aware of the irregularities in the spring of 2006, when a version was submitted to *Nature Physics*, according to sources outside *Nature* familiar with its history. Within days of being notified of the issues, the authors posted a revised version⁴, which added references for all but one image. It also contained a version of the data plot that had been cropped in a way that eliminates the duplicated data points.

In a phone interview, Pennycook and lead author Maria Valera admitted that the data were duplicated on either side of the plot: "I think it was an effort to make it look more attractive to *Nature*

editors, frankly," Pennycook says. "It was the wrong thing to do," adds Valera. She says she was the first to use the mirrored data plot, at an American Physical Society conference, but she does not recall who originally put the plot together. Valera also admits that she did not appropriately cite the earlier paper³ from which several figures were reproduced.

"Obviously mistakes have happened," says Pennycook. "Mistakes always happen in science." He says that he believes the criticisms spring primarily from his rivals at Cornell, and adds that he stands by his work: "A lot of people think we are the number one group in the world," he says. "We have no reason to say anything that's not absolutely right."

The investigating panel of three scientists found "errors in judgement", but no evidence of falsification or fabrication, says panel member Paul Peercy, dean of engineering at the University of Wisconsin at Madison. He says that they concluded that the graph with mirrored data was improperly presented, but that it did not change the paper's core result. "In today's world of PowerPoint," he says, "there is a tendency to present data in whatever the author may think is the best form."

But some in the field continue to have doubts. "There is a pattern of sloppiness here," says Spence. "This is very troubling." ■
 Geoff Brumfiel

1. Browning, N. D., Chisholm, M. F. & Pennycook, S. J. *Nature* **366**, 143–146 (1993).
2. Varela, M. et al. arxiv.org/pdf/cond-mat/0508564v1 (2005).
3. Varela, M. et al. *Solid State Electron.* **47**, 2245–2248 (2003).
4. Varela, M. et al. arxiv.org/pdf/cond-mat/0508564 (2006).
5. Browning, N. D. et al. in *51st Ann. Proc. Microsc. Soc. Am.* 576–577 (San Francisco Press, California, 1993).

Editor's note: In keeping with *Nature*'s practice of confidentiality, *Nature*'s news team was provided with no access to any material pertaining to either of the papers other than the text of the Corrigendum itself.

ORNL/DOE

It's the junk that makes us human

Anyone who has ever put together self-assembly furniture knows that having the right parts is important, but what you do with them can make or break the project. The same seems to be true of the vast amounts of DNA in an organism's genome that used to be labelled as junk. Studies now indicate that this DNA may be responsible for the signals that were crucial for human evolution, directing the various components of our genome to work differently from the way they do in other organisms.

The findings seem to bolster a 30-year-old hypothesis that gene regulation — not the creation of new genes — has moulded the traits that make us unique.

The latest work looks for regions of the genome that have changed rapidly in human evolution, based on the theory that they are most likely to have shaped our differences from other animals. But instead of hunting for rapidly evolving DNA in genes, researchers are starting to look at non-coding DNA — stretches of DNA that don't encode proteins.

In a paper published in *Science* on 3 November, for example, a group led by Edward Rubin of the Lawrence Berkeley National Laboratory (LBNL) in Berkeley, California, reports the result of one hunt through the non-coding genome¹. The team defined a group of 110,549 regions of non-coding DNA that are largely



A. SHAH/NATUREPL.COM

It's all relative: could non-coding DNA be an important factor that separates chimps from humans?

the same in humans and other mammals, reasoning that these regions must be important or they would have changed by randomly mutating over time. The researchers then narrowed the list to 992 regions that have changed markedly in humans compared with

other mammals. Finally, they asked, if these pieces of DNA are regulators, what biological functions do they control?

They found that these stretches of non-coding DNA tend to lie near genes involved in brain-cell function — specifically, in building

Faculty forces president to quit Swiss role

One of Europe's star universities nose-dived into crisis last week when its faculty members forced the president to resign.

Molecular biologist Ernst Hafen, who took over the reins at the Swiss Federal Institute of Technology (ETH) in Zurich last December, submitted his resignation letter on 1 November, a day before a planned faculty vote of confidence he was certain to lose. "It has become clear over the past few days that I don't have the necessary support to continue," Hafen said in a statement.

Hafen had been mandated to reform the ETH by its supervisory council. But his style and his plans both angered and failed to convince senior professors, who banded together to demand that he go.

Hafen is not the first ETH president to have been forced out — there was a quiet resignation in the late 1980s. But the university, which trumpeted its 150th

anniversary last year, has never experienced such hostile internal turmoil.

The ETH came out as the top-performing university in mainland Europe and 27th in the world in the University of Shanghai's 2006 world ranking of universities. It claims 21 Nobel prizewinners, starting with the first laureate, Wilhelm Röntgen, in 1901. Alexander Zehnder, president of the ETH Council, had even greater ambitions: he wanted to raise the university into the top ten worldwide, and asked Hafen to make organizational changes to that end.

Hafen, a highly regarded researcher with a strong publication record, had been considered by the ETH Council — a body nominated for five-year periods by the Swiss federal government — as an ideal candidate to do this. At 50, he was still young. He had experience in setting up a biotech company, The Genetics Company, and he worked in a

research area — systems biology — that is a Swiss priority. Hafen is one of the co-developers of Systems X, the prestigious Swiss national systems-biology network.

Former scientific colleagues refer to Hafen as warm and collegial. But ETH faculty members were clearly not seduced, claiming that his challenges to all aspects of their work encroached on their traditional autonomy. They said that his plans for a rapid reform of the university, a programme known internally as ETH 2020, were ill conceived — and that Hafen often failed to explain which problems they were designed to solve, and how they would solve them. Worse still, critics say, Hafen set up a consultation procedure but then ignored their own carefully formulated input. Attacks on Hafen by some faculty members reportedly became personal.

Key reforms included abolishing the

connections between brain cells. This suggests that the non-coding DNA pieces might orchestrate the wiring of our brains, says team member Shyam Prabhakar, a computational biologist also at LBNL. "This is really very striking," he says. "It's the biggest signal we see in human evolution."

Other genome analysts agree that the study is intriguing, but add a few notes of caution. First, the statistical significance of the link to the brain isn't as strong as some would like. Second, it's not clear whether the rapidly changing DNA is actually changing its function, or just accumulating insignificant mutations — in other words, were the changes caused by natural selection?

But the paper seems to agree with other recent analyses of non-coding DNA. The first, published by David Haussler's group at the University of California, Santa Cruz, used stricter data limits to define 202 rapidly evolving regions of DNA². Many of these were non-coding, but the most dramatically evolving piece of DNA encodes an RNA molecule that is expressed in developing brain tissue³.

Another analysis, not yet published, from Manolis Dermitzakis's group at the Wellcome Trust Sanger Institute in Cambridge, UK, uses slightly different methods to trawl through non-coding DNA and finds 1,500 regions of interest. Some of those overlap with both Haussler's and Rubin's analyses, although

Dermitzakis doesn't find the same association with brain function as does Rubin's group. But in unpublished work Gerton Lunter, a bioinformatician at the University of Oxford, UK, has found a link to the brain when he compares human and mouse non-coding DNA.

Altogether, Lunter says, the various results seem to be defining a set of key non-coding regions. "It all points in the same direction, so I think it's believable," he says.

The challenge, of course, is to figure out how these DNA segments control our fate. On 5 November, Rubin's group published a method for identifying which pieces of non-coding DNA act to enhance the expression of

nearby genes⁴. This is just one possible function of non-coding DNA; elsewhere, it might act as regulatory RNA, or do things researchers can't even guess at yet.

But that's what makes non-coding DNA so much fun, says Haussler. "Here we have a way of discovering new biology," he says. "And looking at that biology is going to be so exciting." ■

Erika Check

1. Prabhakar, S., Noonan, J. P., Pääbo, S. & Rubin, E. M. *Science* **314**, 786 (2006).
2. Pollard, K. S. et al. *PLoS Genet.* doi:10.1371/journal.pgen.0020168; 2006).
3. Pollard, K. S. et al. *Nature* **443**, 167–172 (2006).
4. Pennacchio, L. A. et al. *Nature* doi:10.1038/nature05295; (2006).

"This is really very striking — it's the biggest signal we see in human evolution."



Ernst Hafen failed to gain backing for his reforms.

position of rector, who is elected by the faculty to take charge of university teaching, which under the new plan would be run by the president's office, and changing the ETH's 15 departments into a smaller number of schools to be run by professional deans.

This structure was introduced into Switzerland's other federal university, EPF Lausanne, five years ago and met "a lot of

resistance from the faculty initially", says Patrick Aebischer, the university's president. Aebischer's reforms, which included the addition of a new life-sciences department, are now widely regarded internally and externally as a success.

But at the ETH, Hafen was unable to overcome faculty resistance. As tempers peaked at the end of October, the ETH board, composed of the rector, the president and two vice-presidents, put the ETH 2020 programme on hold — against Hafen's wishes. The heads of all departments except biology, where Hafen remains a faculty member, supported a letter on 24 October demanding his resignation. After attempts at reconciliation and public statements admitting his mistakes, Hafen finally conceded defeat.

The rector, Konrad Osterwalder, will act as president until a new one is appointed, a process that may take at least a year. ■

Alison Abbott

See Editorial, page 123.



COOL MICE LIVE LONGER

The best body temperature for fighting ageing might be 36.5 °C.

www.nature.com/news

Power up your memory bank

Zapping people's brains with an electric current while they sleep might not sound such a good idea. But by boosting natural brain waves it can improve memory, as new work shows. This approach might one day help us learn better or provide new treatments for sleep disorders.

One of the functions of sleep is generally agreed to be the consolidation of memories: certain tasks learned before a snooze are remembered better than if no nap follows. Now Jan Born and his colleagues at the University of Lübeck in Germany think they know why.

One theory holds that levels of brain chemicals, or neuromodulators, are what affects memory during sleep. But Born wanted to test a competing idea — that waves, or oscillations, of electrical activity are responsible. To pinpoint which part of sleep contributes to strengthening memory, the team used a weak electrical current to boost brain activity during deep sleep.

Oscillations in electrical activity occur naturally, at different frequencies, during different sleep phases throughout the night. Oscillations are fastest during rapid eye movement (REM) sleep, and slowest in phases of deep sleep. Boosting the slow waves seems to be the key to boosting memory, says Born.

"By inducing such slow oscillations, we can say for the first time that slow oscillations are the cause" of memory enhancement during sleep, he says. The findings are published online in *Nature* this week (L. Marshall *et al.* *Nature* doi:10.1038/nature05278; 2006; News and Views doi:10.1038/nature05309; 2006).

Some previous studies have suggested a correlation between slow oscillations and memory enhancement. If people learn a task before sleeping, their brains show more slow-wave activity when they are asleep. But other researchers have failed to establish a link, leading some to believe that the waves are functionless by-products of the sleeping brain.

Born and his colleagues had 13 medical students memorize 46 pairs of words before going to sleep. Four electrodes were then strapped to each volunteer's scalp and, once the students were in a phase of slow-wave sleep, the researchers induced a weak electrical current for five periods of five minutes each. The students were allowed to sleep undisturbed until morning, when they were retested on the same word task.

A control group that did not receive brain



Sleep on it: taking a nap might help you remember what you've learnt.

stimulation remembered on average 37.4 words before they went to sleep and 39.5 when they woke up. The test group remembered an average of 36.5 words before sleeping, but recalled 41.2 after the treatment. Born says this improvement is remarkable, because medical students are presumably already skilled at remembering facts. What's more, there was no memory enhancement when the researchers tried to mimic the faster-oscillating brain waves in other phases of sleep, or when they applied the stimulation at the end of the night, rather than the beginning.

The treatment didn't work for all tasks, however. Slowly oscillating waves seem to improve recall of words and facts, but the volunteers were no better at remembering a finger-tapping task. Born suspects that this type of memory is consolidated during REM sleep.

So how does it work? The team suggests that the slow oscillations help the brain replay newly formed memories during sleep, by triggering intracellular signals in nerve cells that strengthen the connections between them. That could explain, Born says, why memory gets better when artificial electrical stimulants help boost the natural brain waves.

But don't expect a do-it-yourself memory-enhancement kit to appear on the market

overnight. It's easy to imagine, says Born, how overnight stimulation might come in handy for cramming for exams or memorizing that big presentation. But the team did not test the volunteers later, to see if the effect persists, and do not know if there are any long-term side effects to the treatment.

"I don't think one can do it at home," says Derk-Jan Dijk, a sleep expert at the University of Surrey's Sleep Research Centre in Guildford, UK. "We still need to find out what potential negative consequences could be if this stimulation is done for a longer period of time. Applying stimulation to the brain is something that for the time being we really would like to do in a controlled environment."

Similar electrically based approaches are sometimes used in psychiatry. Some patients with depression, for example, seem to benefit from a technique called transcranial magnetic stimulation, where an electromagnetic current is generated in the brain by an applied magnetic field.

Born thinks that his ideas too might be used in therapy some day. Perhaps, he suggests, doctors could boost the brain's natural waves to treat sleep disorders, depression or perhaps even the reduction in slow-wave sleep that comes with ageing. But he recognizes these applications are far in the future: "It's just a dream," he says.

Kerri Smith

"Memory gets better when artificial electrical stimulants help boost the natural brain waves."

SPECIAL REPORT

Telling the time

Geochronologists can pin down dates in deep time more accurately than ever before. **Rex Dalton** talks to the researchers who are rewriting the details of Earth's history.

By fine-tuning their techniques, researchers are refining their ability to measure ever more precisely the ticking of Earth's geological clock.

For decades, geologists and palaeontologists have had only ball-park estimates for when major events happened in the history of life on Earth. Now a series of new methods has radically improved their understanding of time long gone¹. With unprecedented precision, researchers are now arguing over whether date estimates are off by as little as 100,000 years — remarkably accurate for events that may have occurred hundreds of millions of years ago.

Leading the quest for increasing accuracy is the international EARTHTIME project, the brainchild of Sam Bowring, a geologist at the Massachusetts Institute of Technology (MIT) in Cambridge. Its goal is nothing less than a complete restructuring of the chronology of deep time². "In a decade, we hope to radically recalibrate Earth's history back to the origin of animals," Bowring says.

Researchers involved in the effort gathered in Philadelphia last month at a meeting of the Geological Society of America to plan out their strategy. Yet as they try to standardize their complex dating procedures, they must accept a re-evaluation of dates that they themselves may have determined.

Last year, Bowring won a US\$1-million grant for research with about 225 collaborators for three years from the National Science Foundation. In January, team members plan to apply for another grant to undertake 'a proof of concept' experiment — seeing whether they can standardize geological dates for the Cretaceous–Tertiary (K/T) mass extinction about 65 million years ago.

The group hopes to date the K/T boundary in a series of sediments in eastern Colorado, and use that date as a benchmark for studies in other geographical regions. They will also target another geologic interval from the Cretaceous period, about 90 million to 100 million years ago, to further cross-check dates and develop a precise chronology.

Douglas Erwin, a palaeobiologist at the

Smithsonian Institution in Washington DC and co-organizer of EARTHTIME, notes that major extinctions are often followed by a burst of speciation. "I think these new dating methods will be a powerful spur to new work on evolutionary rates," he says.

One of EARTHTIME's major thrusts will involve 'astronomical tuning', as Earth's orbital motion affects the geological record. Earth's changing movements — the angle of the globe's spin axis, the path of its orbit and the orientation of its axis relative to the Sun — change cyclically, and in turn influence climate. These climatic changes appear in the sedimentary record as changes in the thickness

of sediment layers, in the ratio of carbon or oxygen isotopes, and in the abundance of tiny fossils, for example.

Scientists have already calculated this astronomical record as far back as nearly 40 million years. Now they plan to push it

back even further.

Traditional techniques to date rocks include identifying fossils that were known to exist during particular eras; counting the ratios of different isotopes of radioactive elements, such as uranium, argon or lead, as these change over time; or relying on the intermittent reversals of Earth's magnetic field, recorded in rocks.



Fossils known to exist at a certain time can help researchers put a date to rocks.



At the rock face: researchers study sediments dating from the Cretaceous–Tertiary boundary.

K. JOHNSON

But although such methods can estimate the age and conditions under which a particular rock was formed, they cannot reveal how long those conditions lasted. That's where astronomical tuning comes in, by providing a long-running sequence of climatic events, such as ice ages, that are recorded in seafloor cores.

"You see variations of colour and lightness in the cores," says Heiko Palike, a geologist at the National Oceanography Centre in Southampton, UK. "It is like counting tree rings, then analysing the isotopic composition of the microfossil shells in the layers."

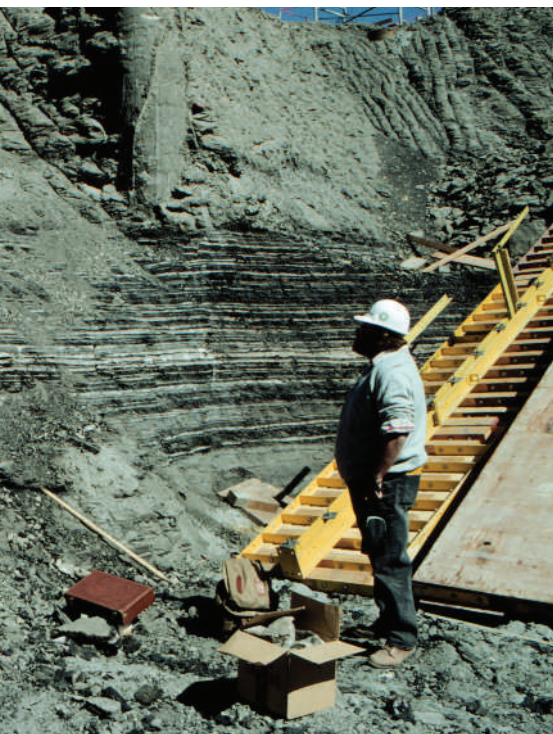
Deep time

To expand this astronomical record, researchers are designing ocean drilling expeditions primarily to examine these time-specific sedimentary cores³. The drill ship *JOIDES Resolution* is set to travel to the Pacific Ocean in November 2007, to drill a core dating between 35 million to 55 million years ago, to help pin down the chronology of events during that time period.

But ocean cores only can reveal sediments going back about 180 million years; any older sediments have been recycled into Earth's interior. When it comes to exploring further back in time, geochemical analyses are the gold standard — specifically, isotopic analyses of radioactive elements as they decay.

Methods include argon/argon dating, which

J. GEHLING/SOUTH AUSTRALIAN MUSEUM



compares the amount of radioactive argon-40 and argon-39, and uranium/lead dating, which does the same with isotopes of uranium and lead. Argon/argon dating can date materials from 2,000 years ago back to Earth's birth, some 4.5 billion years ago. The uranium/lead technique's usefulness ranges backwards in time from about 10 million years ago.

There are some 50 labs worldwide doing argon/argon dating, but of 20 in the United States, only about five will be able to obtain EARTHTIME's desired degree of precision, says Paul Renne, director of the Berkeley Geochronology Center in California.

Researchers use radioactive reagents, typically argon, uranium or lead, to correct for material lost in analysis. But these 'spikes' tend to vary between labs, increasing the margin of error. In a bid to enforce standardization, Bowring bought up all the available spikes, and EARTHTIME investigators now obtain them from Bowring's lab at MIT.

The techniques for uranium/lead dating have also improved. For instance, the Arizona LaserChron Center at the University of Arizona combines a laser with a mass spectrometer to produce uranium/lead dates on zircons — microscopic crystals found in sediments. In the 1990s, multiple zircons often were used for dating analyses, but now a single zircon can be dated accurately, reducing the error from averaging the dates from several zircons.

But the most important development in

uranium/lead dating has been the ability to treat zircons to compensate for the loss of lead from radiation. James Mattinson of the University of California, Santa Barbara, revolutionized the field with a high-temperature technique⁴ for removing zircon surfaces that may have lost lead over geologic time, introducing errors.

The date debate

Mattinson's technique prompted others to reanalyse material from hotly debated points in time, such as the Permian–Triassic boundary (P/Tr), a mass extinction of terrestrial and marine life that occurred some 250 million years ago. The ebb and flow of this dispute — which involves key EARTHTIME members — shows the personal investment researchers have in their dates, and shows how hard it can be for them to jettison their published views.

In 1998, Bowring and colleagues reported uranium/lead zircon dates for the P/Tr boundary of 251.4 million years ago, plus or minus 300,000 years⁵. But in 2001, Roland Mundil of the Berkeley geochronology lab questioned Bowring's date, suggesting that the MIT group had not sufficiently accounted for lead loss in its analysis of multiple zircon batches⁶. After learning of Mattinson's technique, Mundil applied it to P/Tr zircons — and placed the mass extinction at 252.6 million years ago, plus or minus 200,000 years⁷.

The difference between the two figures might not seem much, but both sides clung on to their

view — just as the presses were set to roll for the new edition of a major reference text, *A Geological Time Scale 2004* (ref. 8). The book used MIT's dates, prompting "a great deal of criticism", says James Ogg, a geologist at Purdue University in West Lafayette, Indiana, who co-edited the book. Now, Ogg says, the prevailing view is that Mundil was correct. "We should have listened to Mundil more," he says, and an earlier date for the P/Tr boundary will probably be used in the next edition of the book, scheduled for 2008.

At a conference two months ago in Melbourne, Australia, Bowring reported his team's latest date for the same event: 252 million years, plus or minus 100,000 years. "The philosophy of EARTHTIME is to resolve these kinds of debates to the highest level," he says.

"The philosophy of EARTHTIME is to resolve dating debates to the highest level."

1. Erwin, D. H. *Annu. Rev. Earth Planet. Sci.* **34**, 569–590 (2006).
2. Clarke, T. *Nature* **425**, 550–551 (2003).
3. Cyranoški, D. *Nature* **426**, 492–494 (2003).
4. Mattinson, J. *Chem. Geol.* **220**, 47–66 (2005).
5. Bowring, S. A. *et al. Science* **280**, 1039–1045 (1998).
6. Mundil, R. *et al. Earth Planet. Sci. Lett.* **187**, 131–145 (2001).
7. Mundil, R. *et al. Science* **305**, 1760–1763 (2004).
8. Gradstein, F. M., Ogg, J. G. & Smith, A. G. *A Geological Time Scale 2004* (Cambridge University Press, 2004).

SCORECARD



Property rights

'Air rights' above land have a value to high-rise developers separate from the plot itself. So conceptual artist Jonathon Keats has bought the rights to the higher dimensions predicted in string theory associated with six properties in San Francisco. Although the cost was very reasonable, development opportunities will be limited if the extra dimensions are, as currently predicted, subatomic in scale.



Rewarding virtue

Having been fined in May for despoiling a beautiful lake while making a film, Chinese director Chen Kaige has been nominated for a 'Green Chinese' award. Nominators felt that the fuss over *The Promise* — which is up for a Golden Globe — raised awareness of environmental concerns.

NUMBER CRUNCH

12% of Americans see the environment as one of the three most important issues facing the United States today.

48% see global warming as the most important environmental problem facing the United States, up from 22% in 2003.

45% think there is a lot of disagreement among scientists on global warming, but 61% feel that there is enough evidence to justify action.

60% of Americans would be willing to pay at least \$10 more a month for electricity to solve the global warming problem.

ZOO NEWS

Steve Johnston, of the University of Queensland, Australia, wants to set up a sperm bank for koalas in order to help prevent the spread of sexually transmitted diseases in breeding programmes.



Sources:
Reuters,
MIT, ABC
News

Coffee fortune bails out private university

The most generous private donation ever made to a German university will secure the future of the International University Bremen (IUB). The Swiss-based Jacobs Foundation last week promised to donate €200 million (US\$25 million) to the university, which has faced financial difficulties since it opened in 2001.

Unusually for Germany, the IUB, which offers English-language courses in the sciences and humanities, is privately funded, mostly by private endowments and tuition fees. The windfall from the Jacobs Foundation, a charitable trust founded by the coffee-roaster dynasty Jacobs, will expand the university's research and increase the number of postgraduate researchers and staff from 350 to 600.

But the decision to rename the university 'Jacobs University' has provoked mixed reactions. Some are concerned about scientific independence, but enthusiasts hail the move as the beginning of a more widespread corporate citizenship in Germany's science culture.

Geneticists delete names on the lunatic fringe

Genes with whimsical names that might cause offence to people carrying mutations in them will be rebranded, the committee that adjudicates on such matters has decided. Names such as *lunatic fringe*, *radical fringe*, *Sonic hedgehog* and *Indian hedgehog* will no longer be used to refer to human genes.

A survey by the Gene Nomenclature Committee of the Human Genome Organisation, based at University College London, came up with ten genes that have

"inappropriate, demeaning and pejorative" names, many of which are linked to eponymous developmental defects. Most genes on the list were initially discovered in fruitflies, for which geneticists have a tradition of coming up with jokey names. The human versions will now be known simply by their abbreviations.

Multimillion-dollar endowment for cosmology

Big philanthropic donations to science are becoming more common, but Norwegian-born billionaire Fred Kavli is unusual in funding blue-skies research — in neuroscience, nanotechnology and cosmology.

On 1 November, his foundation announced the last of four new Kavli institutes to be established this year. The Kavli Institute for Cosmology at the University of Cambridge, UK, will get an endowment of US\$7.5 million. This will pay for a new building to bring together researchers from different departments and will support four fellows — but what work they do is up to them.

Other Kavli institutes confirmed in 2006 are at Peking University, the Chinese Academy of Sciences and Harvard University. There is now likely to be a lull as the foundation focuses on the \$1-million dollar prizes it will award for the first time in 2008.

No reward in Japan for alleged bio-spy

Researcher Takashi Okamoto, charged with economic espionage in the United States, was back in court in Japan last week, this time suing the Japanese government for ¥42.9 million (US\$360,000). He lost, but few



Takashi Okamoto leaving detention in 2004.

are likely to feel much sympathy for him.

Okamoto was suing for his 57-day detention during his extradition trial in 2004. He was originally charged with economic espionage in the United States in 2001 for removing reagents from the Cleveland Clinic where he worked before moving back to Japan. He beat an extradition charge in a Japanese court by claiming that extradition cannot be granted when the law broken does not exist in Japan.

At the time, many in Japan felt that Okamoto was in the wrong although the US use of the espionage law was harsh. Okamoto was criticized, however, for not testifying to help his former friend, Hiroaki Serizawa, whose academic research career was destroyed by the episode (see *Nature* 430, 960–961; 2004).

MIT report chastises warring neuroscience units

A report on collaboration within the neuroscience community at the Massachusetts Institute of Technology (MIT) in Cambridge seems to be fuelling discord rather than mending it.

The report was commissioned in July to investigate allegations that Nobel prizewinner Susumu Tonegawa, who heads the Picower Institute for Learning and Memory, inappropriately interfered with the hiring of female scientist Alla Karpova (see *Nature* 442, 341; 2006). In the report, an investigating committee concludes that he did — but says that others involved in the recruitment process did not follow correct procedures and that there is competition and lack of communication between MIT's neuroscience units, particularly the Picower and the McGovern Institute for Brain Research, where Karpova would have been working.

Some MIT neuroscientists are pleased with the report, but others say that it contains major inaccuracies and does not sufficiently reprimand Tonegawa. The MIT administration will establish an advisory council on neuroscience to oversee the warring units.

Acoraceae to Zosteraceae online

A definitive electronic checklist of a large group of flowering plants — the monocotyledons — will be welcome news to green-fingered researchers.

A team from the Royal Botanical Gardens at Kew, UK, publishes the *World Checklist of Monocotyledons* online this week, with a separate downloadable database for the grasses; the two will be integrated by the end of 2007. Its creators hope it will become the first port of call for plant scientists, who until now have had to search for the accepted names and synonyms of their specimens in obscure literature held by specialist libraries. Users can also build customized lists of species by region or country.

Monocots include staple food crops, such as wheat, rice and maize, and many valuable horticultural plants such as lilies, orchids and daffodils. The database holds 75,000 species, representing a quarter of all flowering plants.

► www.rbgekew.org.uk/wcsp/monocots; www.kew.org/data/grasses-syn.html



BUSINESS

A breed apart

The US Food and Drug Administration may soon approve the use of cloned livestock for food. But regulatory roadblocks aren't the only thing keeping clones off the menu, as **Heidi Ledford** reports.

Watson 101 had already lived up to his name by the time he was 12 years old. He had the longest horns of any Texas Longhorn in the world — they spanned 101 inches (260 cm) from tip to tip, and were still growing. Watson could have earned thousands of dollars as a sire, but had been castrated before breeders saw his potential. So Watson's owner, Marquess Arrow Ranch in Texas, decided to order some clones of its prized steer.

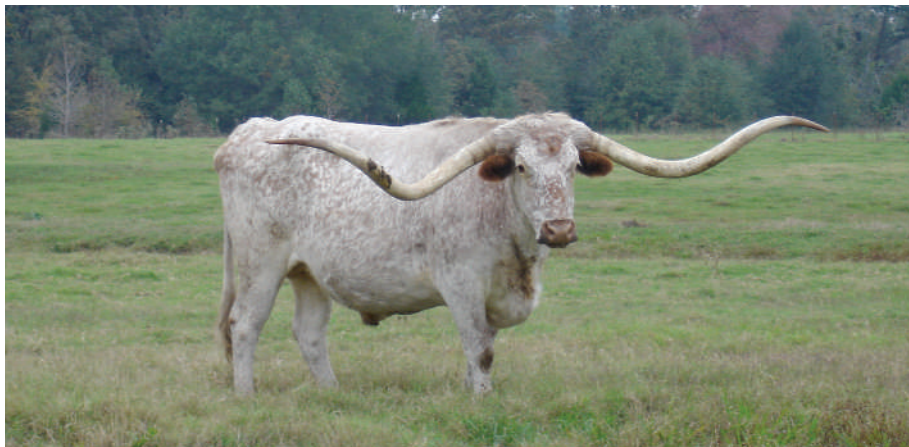
Elite breeders of extraordinary cattle are currently the main customers for a livestock-cloning industry that has struggled to take off since Dolly the sheep, the world's first cloned mammal, was born in 1996. The companies that sprang up to sell the technology have been in limbo as regulators in key markets such as the United States and Japan considered the admissibility of cloned meat into the food supply. "We're chained," says Mal Brandon, chairman of Clone International, an animal-cloning company based in Melbourne, Australia. "Everything's been stalled while we've been waiting for clearance by the US authorities."

The chains are now loosening somewhat. The Food and Drug Administration (FDA) confirmed last month that it will release a draft risk assessment on the matter by the end of the year. But don't expect to find cloned meat in your supermarket in the near future. "The cost of cloning and the lack of efficiency in the process will preclude us from eating it any time soon," says Joe Paschal, a livestock specialist at Texas A&M University in Corpus Christi.

Business interest in livestock cloning is concentrated in the hands of three US companies — ViaGen of Austin, Texas; Trans Ova Genetics of Sioux Center, Iowa; and Cyagra of Elizabethtown, Pennsylvania. Clone International is also active, and there is significant research interest in cloning cattle in Japan.

All these firms have struggled to survive in the absence of regulatory approval. "This area requires quite a lot of cash," says Brandon, "and just to maintain it for research is not good business." Infigen, a Wisconsin-based company that was considered a major player, folded in 2004.

ViaGen president Mark Walton says that his company is not making money and has failed to attract venture capital, relying instead on the



MARQUESS ARROW RANCH

Immortal: Watson 101's genes can live on in his clones.

support of its principal shareholder, billionaire John Sperling. David Faber, president of Trans Ova Genetics, says that it recently attracted some venture capital, but notes that only about 5% of the company is focused on cloning farm animals. Cyagra declined to comment.

And ranchers haven't been stampeding to get hold of clones. Their primary concern is cost. ViaGen sells cloned pigs for around US\$4,000 each, for instance, whereas an uncloned sow might cost about \$150.

This means that the main target for the cloning companies is the top of the sophisticated breeding pyramid that produces today's farm animals. Last year, for example, farmers and ranchers in the United States spent

more than \$225 million on bull semen. Outstanding sires can cost tens of thousands of dollars, says Paschal, with a few elite bulls worth more than half a million dollars. These prices make ViaGen's \$15,000 bull easier to accept.

Even in this niche, cloned animals will make up a tiny proportion of the breeding stock. According to Walton, it takes about 75,000 bulls to replenish the US dairy cattle herd, and 200,000 for the beef cattle herd. Last year, ViaGen produced just 50 cloned cattle.

Part of the challenge is technical. The efficiency of somatic cell nuclear transfer — the technique used to produce Dolly — remains disappointingly low. Faber estimates that the proportion of bovine embryo transfers that result in a live birth at Trans Ova Genetics ranges

from 8% to 25%. And these numbers don't take account of the large proportion of embryos that die before being transferred to a surrogate mother. "The levels of efficiency will keep the number of clones relatively small," says Faber. He nonetheless predicts that demand will increase as efficiencies improve and costs decrease.

Export market

Some companies are trying to build a niche by selling cloned cattle to poorer countries, such as China, to jump-start their cattle-breeding programmes. Xiangzhong Yang, director of the University of Connecticut's Center for Regenerative Biology and founder of Evergen Biotechnologies, claims that a cloned bull could achieve as much on Chinese farms in two years as a breeding programme taking seven years and costing up to a million dollars. And Clone International has already shipped four clones of New Zealand dairy bulls to China for use in research there. But Robert Blake, an animal scientist at Cornell University in Ithaca, New York, cautions that animals bred to succeed in one country may not fare as well when faced with foreign diseases and nutritional standards.

Alison Van Eenennaam, an animal biotechnology specialist at the University of California, Davis, says the hype around animal cloning has exaggerated its likely role in food production. "You hear a lot about it and you think there must be a zillion clones out there," says Van Eenennaam. "But the only place it's really getting used is the upper echelon of the breeding pyramid." ■

"The levels of efficiency will keep the number of clones relatively small."

— David Faber

DUTCH COURAGE

Most astronomers head for remote mountain-tops or deserts to study the cosmos. **Jenny Hogan** meets a confident team set up on a patch of farmland in a crowded corner of mainland Europe.

In the lobby of Astron, the Netherlands Foundation for Research in Astronomy, a clear plastic column almost full of dirt commemorates an unusual aspect of the institution's struggle to study the cosmos. Over the past three years, the institute's astronomers have been negotiating with 42 Dutch farmers to buy up the 400 hectares of land they need for the first stage of a remarkable new radio telescope — the low frequency array, or LOFAR, a €148-million (US\$188-million) attempt to open up a new part of the spectrum to astronomical inspection. Every time a stretch of land was acquired, a bit of soil was added to the column in the lobby, making a layer cake of peaty soil and sand.

Out in the fields, the business end of LOFAR consists of sites dotted with low-slung, leggy wire antennas liberally spattered with the droppings of the birds that perch on them and a few grey metal Portaloo-like cabins. But the fields aren't the right place to look for LOFAR's grandeur: the telescope's impressive heart is being built in cyberspace. It is there that the faint noise of the Universe may eventually be separated from the background roar of Holland's air-waves and the Milky Way.

'Software telescopes', which couple a vast capacity for absorbing information with a matched capacity for analysing and reassembling it, are widely seen as the future of radio astronomy. Such technologies will be crucial to the Square Kilometre Array (SKA) with which the world's radio astronomers hope to revolutionize their field by the 2010s. At LOFAR, though, the technology faces an extra problem. To secure funding, the array's core had to be built in the Netherlands. Nowhere on Earth is completely free of radio signals, but the Netherlands, the most densely populated country in one of the most built-up regions of the world, is quite remarkably noisy.

To tune in to the early history of the Universe, LOFAR must studiously ignore TV shows, FM radio stations, air traffic control and local taxi drivers, not to mention purely accidental emissions. Heino Falcke, LOFAR's international project scientist, gestures to a tractor trundling past one of what will be 77 fields of antennas. "We would probably measure that as a series of sparks going past our telescope." LOFAR's rivals, cloistered in a remote part of China and the Australian desert, enjoy far quieter environments. Whether LOFAR's software cleverness and number-crunching prow-



ess will allow it to overcome the disadvantage remains to be seen.

Get close enough to the 'Portaloos' in the field and you can hear them hum. Unlock their doors and you find tangles of cable and racks of electronics and blinking lights. This is where the wire roots of the antennas end up, and where the software telescope begins to take shape.

Although the antennas may look simple, they can produce data at an alarming rate. One antenna site, consisting of 96 spindly antennas collecting low-frequency waves and 1,536 squat structures detecting higher frequencies, will produce about 500 gigabits of data every second — far more than even the most sophisticated optical telescopes.

Cyber optics

The humming Portaloos are there to make sure that most of the data are left in the field, reducing a flood to a mere torrent. But that torrent still outdoes the data rate for the detectors on the LHC, the vast particle collider currently being completed at the European Centre for Particle Physics (CERN) near Geneva — instruments whose vast data flows have led to a world-wide effort in GRID computing. The torrent from the fields, channelled through optical fibres, ends up in a windowless strip-lit room at the University of Groningen. Here, an IBM Blue Gene/L supercomputer called Stella compares the data from the 50,000 channels measured at each station with the equivalent data from all of the other 76.

At Astron, they think that this 34-trillion-operations-a-second wonder could restore them to the glories of the 1950s, when the 25-metre Dwingeloo dish, now rusting away behind their institute, was briefly the largest steerable radio telescope in the world. In the 1990s, the institute decided to follow a path that concentrated less on dishes and more on digital. "We thought that Moore's law would be



PHOTOS BY JENNY HOGAN/ILLUSTRATION BY DANNY ALLISON

a good thing to hook onto,” recalls Harvey Butcher, Astron’s director since 1991. To maximize the effort that could be put into data processing meant that the cost of picking up the signals in the first place had to be minimized, leading to the use of cheap dipole antennas.

LOFAR’s dipoles pick up long-wavelength signals at frequencies from 30 to 240 MHz. This long-wavelength radio is a little-explored part of the spectrum, not only because it is widely used on Earth, but also because it is affected by the ionosphere — another source of noise that has to be removed with the help of computers. What’s more, the long-wavelength sky is filled with the synchrotron emission given off by electrons being whipped around by magnetic fields in the Galaxy, the glare from which masks intriguing cosmological signals some 10,000 times as faint.

In 2001, the Astron team calculated the computing requirements for a long-wavelength system big enough to find those faint signals. Then they had a look at the Top 500 list that ranks the world’s quickest supercomputers. It looked as though, given the way that performance was improving over time, the power they needed might be purchasable in the near future. And so it was. The Blue Gene/L in Groningen is three times as powerful as the most powerful machine in the world was back then — and doesn’t quite make it into the Top 500’s top ten today.

Perhaps the most interesting of the feeble signals that the computer will be sifting out are those emitted in the dark ages of the first few hundred million years after the Big Bang, a chapter in the Universe’s history that remains obstinately closed. The dark ages’ darkness stems from the fact that the glowing plasma of the Big Bang, details of which are still visible in the cosmic microwave background, had cooled down into neutral atomic hydrogen. It was only after the stars and galaxies had come to life that their light burnt off this hydro-

“I don’t think you can do the measurement there.”

— Jacqueline Hewitt

gen mist, reionizing the gas into a transparent plasma.

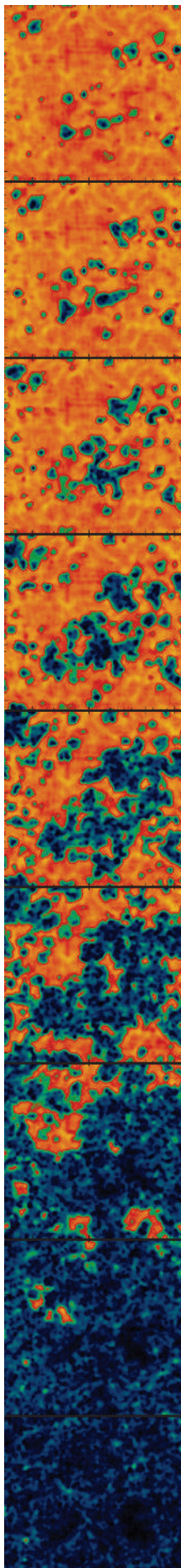
Huge sums of money are being spent on various telescopes that promise to peer back at this epoch of reionization. The \$650-million Atacama Large Millimetre Array, under construction in the Chilean Andes, aims to detect the infrared glow of hot dust around the very first galaxies, radiation that will have been stretched, or ‘redshifted’, by the expansion of the Universe to millimetre wavelengths. NASA’s \$5-billion James Webb Space Telescope, scheduled for launch in 2013, will look for the ultraviolet light of these first galaxies, now redshifted into the infrared wavelengths that the telescope is optimized for.

Shedding light

LOFAR’s approach, shared by the Mileura Widefield Array (MWA), planned for the outback of western Australia, and the 21CMA installation in China, is to look at the neutral hydrogen itself. Neutral hydrogen emits radiation with a wavelength of 21 cm — hence the name of the Chinese array. That 21-cm radiation from the epoch of reionization is redshifted down into the metre-long wavelengths the new telescopes can detect.

What the arrays will actually be looking for are the holes in the 21-cm radiation caused as the reionization centred on stars and galaxies took hold, creating a Swiss-cheese effect. In principle, different stages in the growth of these holes could be teased out by finding the 21-cm hydrogen line at different redshifts — the longer the wavelength, the older the radiation. By moving up and down the frequency dial, it would be possible to watch these bubbles expanding and merging as the Universe opened itself up to inspection. Simulations of the early Universe produce beautiful pictures of this process.

But images from this first generation of telescopes will



A nine-step simulation shows 21-cm emissions being eaten away as reionization spreads through the young Universe.

be, at best, very fuzzy. The head of LOFAR's reionization project, Ger de Bruyn, demonstrates what might be seen by bringing up the theorist's beautiful simulations on his laptop and then removing his glasses, scooting his chair to the far side of his office, and squinting: "This is more likely." The first detection of reionization is more likely to be statistical, giving a measurement of how the brightness of the glow varied from one time to another, than a resolvable picture of specific bubbles. And even a statistical measurement is going to be tricky. Xiang-Ping Wu, chief scientist for 21CMA, says that it could take five years to be confident of seeing a reionization signal.

With 21CMA already taking data, construction on LOFAR started this summer and the first antennas of Australia's MWA expected to be installed next year, it could, says Butcher, "be a really interesting race". The Chinese array, built for just US\$3 million, is the smallest and cheapest. It is located in the west of the country, in a remote valley in Xin Jiang province. When the project was approved in August 2004, "we pitched a tent in the Ulaistai valley and started the construction on the same day", says Wu. The completed array consists of 10,287 log-periodic dipoles — television antennas, more or less — arranged along two perpendicular arms roughly four and six kilometres long.

Unlike LOFAR and the MWA, which have other scientific fish to fry, 21CMA is essentially interested only in the reionization signal. Its field of view is restricted to a patch of sky around the North Celestial Pole and its resolution rather coarse. Although it has been collecting data for three months, no money has yet been allocated to continue operations. But Wu isn't worried. The Chinese Academy of Sciences and National Astronomical Observatories of China, which funded the array's construction, are likely to continue to support the project, he says, and concentrating on a small patch speeds up data acquisition.

Universal challenge

Close behind is LOFAR, which will use two types of antenna, one gathering low-frequency radiation, between 30 and 80 MHz, the other higher frequencies, between 110 and 240 MHz. Because of its geographical situation, LOFAR has a whole suite of systems to deal with radio interference. It will use software filters to flag and discard signals that are too strong to be coming from space. Some noise will be fleeting, and so average out over a long observing time, or be at a single frequency that can be filtered out. This will leave gaps in the data, but the team estimates that they can use about 90% of the spectrum at any one time. The exception is the commercial FM band, which LOFAR will simply discard. This is why there's a gap between the ranges of the two antennas. And that could pose a problem for the study of reionization.

Current models predict reionization would have been getting into its stride by the time the Universe was 500 million years old. If they're correct, "the frequency will be right in the middle of the LOFAR band. Then LOFAR

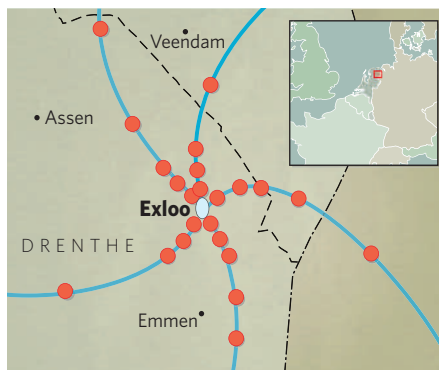
can really have a go at it," says de Bruyn. But if galaxies formed a bit more quickly than predicted, the hydrogen would have burnt up a few million years earlier. In that case, its redshifted signal will lie in the impenetrable realm of Exxact FM and Radio Veronica.

Three years ago this blank spot, and other concerns about the effects of noise, led to a split in the international collaboration of astronomers planning the telescope. The US and Australian members left after the Astron team and their Dutch collaborators committed themselves to a site in the Netherlands. Some of the Americans opted for a Long Wavelength Array in New Mexico, which is not pursuing the 21-cm signal from reionization, the others for the MWA in Australia. "We didn't agree with the decision to build in northern Europe. We didn't think it was wise," says Jacqueline Hewitt, director of the Kavli Center for Astrophysics and Space Research at Massachusetts Institute of Technology in Cambridge, now part of the collaboration working on the MWA in Australia. "I don't personally think you can do the epoch of reionization measurement there — I might be wrong, but I wasn't prepared to spend the next ten years of my life trying."

The site chosen for the MWA — a 12-hour truck drive from Perth, the nearest big city — is beset by very little interference.

But the Dutch part of the collaboration was tied to building their telescope in the Netherlands by a generous dollop of funding. LOFAR was granted €52 million in 2003 from the national Fund for Economic Structure, which reinvests money made from sales of the Netherlands' natural gas. This money has to be used to strengthen the national economy, and so couldn't be spent on building a telescope in another country.

The Dutch contingent wasn't going to walk away from this kind of cash (see 'Making it pay').



Antennas of the the Dutch LOFAR telescope will be centred on the town of Exloo.

Space race

The MWA sees itself more as an experiment — and as a forerunner to the SKA — than as a dedicated piece of infrastructure for many users, which is what LOFAR aspires to be. It has fewer antennas, of only one type, all in one place, and a much smaller project team relying on postdocs and graduate students for a lot of labour. And there are far fewer farmers to negotiate with. This makes it much cheaper than LOFAR, with an expected initial cost of just US\$10 million — although Rachel Webster, a member of the MWA team at the University of Melbourne, says she expects the team will actually spend much more. The initial money is being provided by the US National Science Foundation, the Australian Research Council and the government of Western Australia, among others, but the project is still at the prototype stage. The first 32 of its 500 antenna elements are due to be installed in the outback in 2007.

Hewitt notes that, regardless of who sees the reionization signal first, it will be good for the measurement to be repeated. "We're going to need to compare our results. If one group sees it, I don't think people will believe it until another group has too." Butcher, while regretting the break-up of the project team, is matter-of-fact about its consequences. "On the one hand, there are some bits of the spectrum that we can't use. On the other hand,

Making it pay

In addition to receiving €52 million (US\$66 million) from the Dutch government, Astron's new LOFAR telescope also got €22 million from the region of Drenthe — “a huge amount of money to invest in something as weird as astronomy,” says Astron's director Harvey Butcher.

LOFAR won both investments because it promised to contribute to the Dutch economy.

When teams laid the optical fibre that will carry the telescope's data to a remote processing centre, they also installed fibres to villages in the region that didn't have broadband Internet access.

Other projects are capitalizing on LOFAR's infrastructure for data gathering. “We provide a USB port in these fields. You hook your sensor in,” explains Eugene de Geus,

LOFAR's managing director. There are plans for networks of seismometers, infrasound detectors, weather sensors to provide microclimate data for farmers and wind sensors to improve the models used by northern Europe's wind farms. “Every audience we talk to thinks of something different we could do,” says Butcher.



D. GARDENIER

we have a telescope and they don't.”

LOFAR and the MWA will use their all-sky views to study more than just the epoch of reionization. “The epoch of reionization is the lighthouse. It is the project that most people are interested in,” says Falcke. But the fact that they will be looking through a new window of the spectrum offers a prospect that he sees as just as tantalizing. “There's at least a chance that there is something out there that we haven't found yet.”

Estimates based on the number of objects detected in surveys at higher frequencies suggest that LOFAR's surveys will turn up some 100 million radio-emitting objects. The view at the lowest frequencies should include bright radio galaxies more distant than any seen before. The data stored in LOFAR's buffers will let it zoom in on transient flashes in the radio sky, which could be the radio counterparts of poorly understood γ -ray bursts, or even bursts of radio emission from extrasolar planets, like those that come from Jupiter. Searches for transient events will ‘piggyback’ on all the other projects, commandeering the telescope if something particularly spectacular shows up.

Long wavelengths also have implications for the search for extraterrestrial intelligence — a topic that stirred up significant interest among the farmers selling Astron their land. Working at wavelengths where the earthly background noise makes things difficult becomes a plus when looking for radiation leaking out from civilizations around other stars — at least it does if they have FM radio stations and the like¹.

Network building

Closer to home, flashes of lightning and radio showers from cosmic rays will also be picked up by the array. Falcke's pet project looks at both of these, testing theoretical suggestions that cosmic rays might sometimes trigger lightning. A prototype LOFAR station in Germany, called LOPES, has already managed to image cosmic-ray showers with a time resolution of 30 nanoseconds².

The sharp resolution needed to study individual sources such as galaxies and even planets makes the LOFAR team keen to spread their antennas beyond Holland's borders; whereas the core set of elements doing the reionization studies will be clumped near the Dutch town of Exloo (see map), other antenna elements could be hundreds of kilometres away. Universities in Germany have expressed interest in at least six stations, and a UK group is interested in three. There are also discussions going on with astronomers in Sweden, France and Italy. Falcke hopes that, as the project gets under way, a me-too effect will lead to interest snow-

“We have a telescope and they don't.”
— Harvey Butcher

balling and the array spreading further, expanding from the initial 77 stations to more than 100.

Long baselines, though, exacerbate a problem of calibration. Incoming radio waves are distorted on their way through Earth's ionosphere. LOFAR will have to correct the signal from every station separately, using the results from a model of the ionosphere running on a Linux cluster. This model will be updated every ten seconds by comparing the apparent positions of 200 sources on the sky with their known positions, with the results being fed into the Blue Gene/L. There are now so many cables plugging into the computer's vitals that the doors of its case barely close.

Nor do the challenges end with the inputs. While the various filters and calculations compress the immense amount of data the system gathers, the Blue Gene/L will still be outputting correlated data at an astonishing 39 gigabits per second. That can't be stored for long in its unprocessed form, and the team plans to keep the data for only a week or so. After that, the data will only be available in the form of star maps from which quite a lot will have been lost. Astronomers who want to preserve the primary correlated data will have to take them away while they are still fresh, either through high-capacity fibre or saved to high-density storage media that can be shipped.

And LOFAR is relatively simple compared to the demands of the SKA, with its square kilometre of collecting area and far greater frequency range. When Kjeld van der Schaaf, who is overseeing development of LOFAR's computer systems, makes presentations on the SKA, he includes a slide in which the trends of the Top 500 computer list are extrapolated through to 2015. The SKA's correlating system, which needs to be 3,000 times faster than LOFAR's Blue Gene/L, sits at the top of this ranking.

While the participants in the SKA wait for computing power to catch up with them, design their technology, plead for their funding and choose their site (it will be a radio-quiet spot in South Africa or Australia), LOFAR will forge ahead with its ground-breaking work. And although winning the race to detect the 21-cm signal from reionization is not the project's be-all and end-all, it does have a certain sentimental significance. The very existence of the 21-cm hydrogen line was first predicted by a Dutch astronomer in 1944, but US astronomers beat his compatriots in the race to find it. This time round, they want to get there first. ■

Jenny Hogan is a reporter based in Nature's London office.

1. Loeb, A. & Zaldarriaga, M. <http://arxiv.org/abs/astro-ph/0610377> (2006).
2. Falcke, H. et al. *Nature* **435**, 313–316 (2005).

DOES IT HURT?

Working out whether premature babies feel pain has important implications for child development, says **Jane Qiu**.



It was a precarious beginning. Teresa was born 14 weeks earlier than she should have been. Having spent 26 weeks in the womb, she now lies motionless in an incubator in the neonatal intensive-care unit, relying on mechanical ventilation for breathing and intravenous tubes for nutrition. Her tiny, fragile body is covered with sensors to monitor blood pressure, heart rate, breathing and temperature.

Teresa's destiny is far from certain. She is suffering from the common complications associated with premature birth, such as breathing difficulties, anaemia and infections. She has to go through many routine medical procedures every day, and also faces a number of major operations in the future.

Teresa is a fictional but representative composite of hundreds of thousands of premature babies born each year. Thanks to medical and technological progress, babies born after 23 weeks' gestation in developed countries are now routinely kept alive. But many of them have life-threatening complications. It is estimated that each baby in a neonatal intensive-care unit is subjected to an average of 14 procedures a day — a number that can go up to 50 in extreme cases¹. Although clinicians and nurses intuitively feel that most of these procedures may be painful, there is no easy way to discover whether the babies are feeling anything.

It's a vexed question. Medics rely on self-reporting for pain assessment, but babies can't

use words to explain exactly what they are feeling. For a long time, many health professionals assumed that newborns, and those born prematurely in particular, were too young to perceive pain in the same way that older children and adults do — or if they did, were not able to remember it. They were wary of giving painkillers to infants 'just in case', as the effects and dosing of these drugs had been tested only in adults. Some neonatal clinics started to give painkillers to premature babies during certain medical procedures in the late 1980s but, as

yet, there is no consensus on best practice.

"Most intensive-care units have no guidelines for managing pain in preterm babies, so it's all very patchy and up to the individual clinicians and nurses on duty," says Judith Meek, a paediatric consultant at the Elizabeth Garrett Anderson and Obstetric Hospital in London.

The question of whether preterm babies feel pain goes beyond the obvious ethical and humane issues. In recent years, neuroscientists have unearthed evidence suggesting that the nervous systems of premature babies who experience repeated procedures that are potentially painful develop abnormally. Such children may be either excessively susceptible or desensitized to pain as they get older. The problem is likely to increase in the future, as the number of premature births has been rising in westernized countries². This is partly because of the increased number of multiple

births resulting from more widespread use of fertility treatments.

There are further implications for neuroscientists too. They hope that, as well as answering the question of whether premature babies feel pain, their results will also help them understand more about how an infant's nervous system develops. This could form the groundwork for the development of much-needed drugs specifically designed for children, whose physiology is often different from that of adults.

Pain barrier

At the moment, practitioners use 'pain-coding scales' to assess the amount of pain babies might feel. Each scale consists of a set of behavioural criteria, such as crying, withdrawal and facial expressions, as well as physiological indicators, such as heart rate, blood pressure and breathing patterns. "All those measures have their strength," says Maria Fitzgerald, a neuroscientist at University College London. "But they may not be true reflections of pain experience, as they can also change under other stressful conditions, such as hunger and cold."

Part of the problem is that the experience of pain is far more than just a straightforward physiological response to stimuli. Being a psychological state, pain is always subjective, and is coloured by a number of cognitive and emotional factors. It can result from nociception — the physiological detection of noxious stimuli by pain receptors and nerves — but can also occur independently of this, as a sufferer of phantom-limb pain will attest.

Animal studies suggest that nociception is

"Most intensive-care units don't have guidelines. It's all very patchy and up to the clinicians and nurses on duty."
— Judith Meek

K. COOPER/ARKANSAS CHILDREN'S HOSPITAL

likely to occur in premature infants³. But to work out whether premature babies experience pain, neuroscientists need to tackle two key questions. The first is whether the signals generated by the stimuli actually reach the cerebral cortex, the part of the brain that is crucial for the perception of bodily sensation. The second is: to what extent are premature infants conscious, that is, sufficiently self-aware to experience and be emotionally distressed by pain? Have they yet developed minds?

The debate over whether pain signals reach the cortex in premature babies was fuelled by work first published in 2002 by Tim Oberlander, a paediatrician at the University of British Columbia in Vancouver, Canada. He showed that premature babies with severe brain injuries that disable much of the cortex have the same facial expressions and behaviours in response to potentially painful stimuli as normal babies of the same age⁴. This suggested that the behavioural and physiological responses of premature babies to such stimuli are merely reflexes put together at the level of the spinal cord and brain stem. If signals of pain do not go beyond the brain stem, it is unlikely that preterm infants can feel pain.

Recently, however, a number of scientists have attempted to see inside the brains of premature babies to find out if this is the case. Two teams of researchers, one led by Fitzgerald, the other by Kanwaljeet Anand at the University of Arkansas for Medical Sciences in Little Rock, have independently shown that the cortex of premature babies is activated in response to potentially painful stimuli^{5,6}. Both groups used near-infrared spectroscopy (NIRS), a non-invasive imaging technique that can measure subtle changes in the concentration of the blood's oxygen-carrying molecule haemoglobin.

Neuronal activity in the cortex is coupled with changes in the blood flow to different regions of

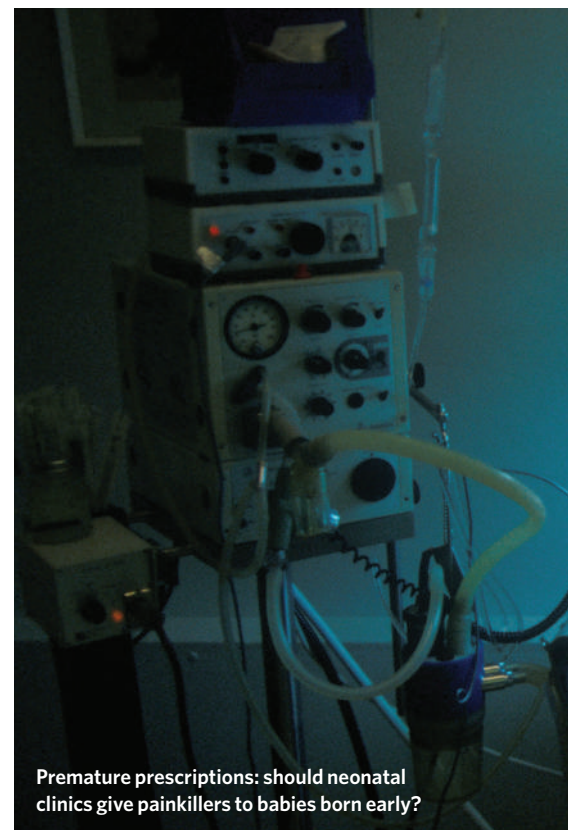
the brain, and this is reflected by the ratio of oxygenated haemoglobin to its deoxygenated counterpart. Because of this, NIRS can be used to monitor neuronal activity in the cortex. It is not as sharp or accurate as other imaging systems, such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET). But these systems would involve moving a very sick and fragile infant into a noisy and stressful environment. NIRS can be performed by simply placing sensors on a baby in its cot.

Using this method, Fitzgerald's team measured cortical activity in response to a heel lance, a widely used procedure for collecting small blood samples, in 18 babies of between 25 and 45 weeks of gestational age, and postnatal ages ranging from 5 to 134 days⁵. Unexpectedly, even in the youngest babies, there was a small but very clear evoked activity in the somatosensory cortex, the part of the cortex that processes bodily sensation.

A brief, non-invasive mechanical stimulation of the heel did not produce detectable cortical activation, even when babies withdrew their feet in reflex action. The researchers found that the older the baby, the bigger the response in the cortex.

Conscious thoughts

In a parallel study, Anand and his team analysed 40 preterm babies, born at 28 to 36 weeks of gestation, at 25 to 42 hours after birth⁶. They used vein puncture, another way of collecting blood samples, to generate potentially painful stimuli. Following either tactile (skin disinfection) or painful (vein puncture) stimuli, there was an increase in blood flow to the somatosensory cortex, but not to other areas of the cortex. The increase induced by vein puncture was much larger than that by skin disinfection, supporting the findings made by Fitzgerald's team. But unlike Fitzgerald, Anand found that



Premature prescriptions: should neonatal clinics give painkillers to babies born early?

the more premature the baby at birth, the bigger the response to painful stimuli.

This difference between the two studies may be explained by the fact that the teams used different stimuli and babies of different ages. Even so, they do have the same central finding in common: there is activation of the cortex following acute pain in premature babies. The finding is met with great enthusiasm from other pain researchers.

"To make this kind of breakthrough in human neonates is extremely difficult," says Ruth Grunau, a paediatric psychologist also at the University of British Columbia. "It presents a major step forward in the field." But proof of cortical activation in babies is not exactly the same as proof of consciousness.

So how sure are the researchers that those preterm infants do have the same subjective experience that adults recognize as pain? There is simply no proof of consciousness, even in wide-awake adults. And recent studies have raised the question of whether comatose adults, who are also unable to speak, might have some level of consciousness, although this remains highly controversial⁷. So proving what a baby feels is currently impossible. "How can you take any biochemical or biophysical measures and equate that to a subjective, conscious experience?" asks Anand. "It is a conundrum."

Anand thinks that the best way forward, for now, is to infer that premature babies can feel pain. "If a baby manifests the same behavioural and physiological responses to pain as seen in adults, together with activation of the cortex, then we can be fairly confident that he is feeling pain, even though that's no proof of conscious-



Instant reaction: behavioural responses to pain could be merely reflexes in the nervous system.



H. SYKES/ALAMY

ness,” Grunau agrees. “We would seem to be holding an extraordinary standard if we didn’t infer pain from all those measures.”

That being the case, the issue of alleviating pain and suffering becomes paramount, especially now that scientists know that repeated painful experiences can alter the development of the nervous system. Some studies show that children born prematurely become less sensitive to pain, whereas others demonstrate that children, especially those who were born prematurely and who have had multiple operations early in life, need a higher dose of analgesics when they need further surgery⁸. In other words, early repetitive pain experiences can either increase or dampen sensitivity to pain later on in life. “There are a lot of confounds here,” Fitzgerald acknowledges. This is not surprising, given the complexity of the phenomenon and the difficulties with human neonate research.

Subjective responses

Studies with newborn rats also show that, depending on the timing, nature and duration of the painful stimuli, the animals become either under- or over-sensitive to pain later on. Both are potentially problematic: children who are less sensitive to pain may be more prone to injury, whereas those who are hypersensitive may withdraw from everyday activities.

Grunau is also concerned about changes in such children’s stress responses. She and her colleagues have conducted detailed studies to correlate the amount of potentially painful procedures premature babies undergo in intensive-care units with their levels of stress

hormones at various stages of early postnatal development. They found that premature babies of 8 and 18 months have much higher levels of stress hormones than their full-term counterparts^{9,10}, although how long these effects persist remains unknown.

A large body of animal and human studies has shown that prolonged exposure to stress hormones can result in changes in the brain, especially in areas important for learning and memory¹¹. “Preterm babies are vulnerable because the structure of their brains is still being formed,” says Meek. “This is really worrying.”

Detailed analysis of a large number of children between the ages of 5 and 14 indicates that children who were born prematurely are about 2.6 times as likely to have attention-deficit disorder as their full-term equivalents¹². Such children also tend to have various learning difficulties, increased levels of anxiety and poorer cognitive outcome. It is difficult to pinpoint the precise causes of these differences — after all, prematurity brings a host of problems that could all contribute in one way or another. But, says Grunau, pain is certainly on this list of potential factors.

So researchers are trying to develop ways to measure pain more accurately in premature babies. Fitzgerald is now testing how well cortical activity in response to potentially painful stimuli correlates with subjective assessment of pain by medical practitioners. She is also using imaging to monitor the efficacy of painkillers

in the babies. To refine the findings with NIRS, her team is now using electroencephalography, which will allow the group to map how input from nerves is transmitted and processed in various regions of the preterm baby’s cortex. Anand and his colleagues are carrying out similar kinds of studies including the effects of prolonged pain, such as post-operative pain and discomfort from being on a ventilator.

Anand’s team is also using fMRI and other imaging techniques to broach the highly controversial and politically sensitive question of whether the fetus can feel pain. Researchers agree that studies on premature infants won’t address this question. The process of being born alters an infant’s brain physiology so radically that it cannot readily be compared to that of a fetus still in the womb. Even with the help of sophisticated imaging technology, tackling this question will be far from easy.

Studies of pain in premature babies could, however, fuel an initiative to develop painkillers designed especially for babies and children. Drug regulatory bodies, government funding agencies and charities have now realized this is an important issue.

In the United States, the Best Pharmaceuticals for Children Act of 2002 authorizes government agencies, in particular the National Institutes of Health and the Food and Drug Administration, to help develop drugs specifically for children. Task forces of scientists, clinicians and ethicists specialize in various conditions affecting newborns, including pain as well as neurological, cardiovascular and respiratory diseases. The initiatives aim to set priorities for scientific research and to establish ethical guidelines for clinical trials using babies.

One such task force, the Neonatal Pain-Control Group, “is one of the top priorities among all the task forces,” says Anand, its chairman. “We expect to see some major clinical advances in neonatal pain management

within the next few years.”

Jane Qiu is a science writer based in London

“How can you equate any biochemical or biophysical measures to a subjective, conscious experience.”
— Kanwaljeet Anand

1. Simons, S. H. P. *et al. Arch. Pediatr. Adolesc. Med.* **157**, 1058–1064 (2003).
2. Kramer, M. S. *et al. J. Am. Med. Assoc.* **280**, 1849–1854 (1998).
3. Fitzgerald, M. *Nature Rev. Neurosci.* **6**, 507–520 (2005).
4. Oberlander, T. F. *et al. Pediatrics* **110**, 570–576 (2002).
5. Slater, R. *et al. J. Neurosci.* **26**, 3662–3666 (2006).
6. Bartocci, M. *et al. Pain* **122**, 109–117 (2006).
7. Hopkin, M. *news@nature* <http://www.nature.com/news/2006/060904/full/060904-11.html> (2006).
8. Peters, J. W. *et al. Pain* **114**, 444–454 (2005).
9. Grunau, R. V. E. *et al. Pain* **58**, 341–346 (1994).
10. Grunau, R. V. E. *et al. Sem. Fet. Neonat. Med.* **11**, 268–275 (2006).
11. Kloet, E. R., Joels, M. & Holsboer, F. *Nature Rev. Neurosci.* **6**, 463–475 (2005).
12. Bhutta, A. T. *et al. J. Am. Med. Assoc.* **288**, 728–737 (2002).

An open letter to Colonel Muammar al-Gaddafi

Dear Colonel Muammar al-Gaddafi,

We, Nobel Laureates in the sciences, are gravely concerned about the ongoing trial of five Bulgarian nurses, Valya Cherveniyashka, Snezhana Dimitrova, Nasya Nenova, Valentina Siropulo, Kristiana Valcheva, and a Palestinian doctor, Ashraf Ahmad Jum'a, in Tripoli. The six face death-penalty charges of deliberately infecting 426 children with HIV at al-Fateh Children's Hospital in Benghazi in 1998. Strong scientific evidence is needed to establish the cause of this infection. However, independent science-based evidence from international experts has so far not been permitted in court.

Libya is currently making efforts to join the community of peaceful nations by renouncing weapons of mass destruction and adhering to international standards regarding the rule of law. This trial is another opportunity for Libya to demonstrate its commitment to recognized values and norms. But so far Libya has failed to follow the norms of international justice in the case of the charged medical workers.

We appreciate the agony and the sadness

of the parents of these children and we sympathize with the difficult situation of the Libyan authorities in trying to deal with this matter. However, we feel that if justice is to be served it is essential that the defence should be permitted to present its case.

Among the disallowed scientific evidence is a 2003 report, which Libya requested, and which was provided by Luc Montagnier, a co-discoverer of the virus that causes AIDS, and Italian microbiologist Vittorio Colizzi. The report concluded that the infection at the hospital resulted from poor hygiene and reuse of syringes, and also that the infections began before the arrival of the nurses and doctor in 1998.

On 29 August 2006, a Libyan prosecutor reiterated the call for the six to be given the death penalty. The next, and probably last, court hearing is scheduled for the 4 November, with a verdict expected shortly thereafter. A miscarriage of justice will take place without proper consideration of scientific evidence. We urge the appropriate authorities to take the necessary steps to permit such evidence to be used in this case.

To uphold justice, and ensure a fair trial, we affirm the need for:

- Defence lawyers to have the right to call and examine witnesses on the health workers' behalf under the same conditions as witnesses called against them, and
- The appropriate authorities to call upon internationally recognized experts in AIDS research to examine and testify on the evidence as to the cause of the HIV infections in the children.

Yours sincerely,

Richard J. Roberts* and 113 fellow Nobel Laureates†

*1993 Nobel Laureate in Physiology or Medicine, Chief scientific officer, New England Biolabs, 240 County Road, Ipswich, MA 01938-2723, USA

†A full list of signatories to this letter is available as supplementary information at www.nature.com/nature/journal/v444/n7116/supinfo/444146a.html.

Published online 2 November 2006.

Don't forget the steps that led physics to where it is

SIR — In your News story “Neutrinos make a splash in Italy” (*Nature* **443**, 126; 2006), you highlight a project, “first sketched out 25 years ago”, to attempt to discover whether neutrinos have mass.

As president of the National Institute for Nuclear Physics from 1977 to 1983, what I presented to the Italian authorities 27 years ago, far from being a ‘sketch’, was a full project. It had a set of scientific goals, including the neutrino oscillations from the ‘artificial’ source located at the CERN particle-physics lab, where, in the early 1960s, the search for the third lepton was intensively carried out using the hypothesis of the third neutrino, now called ‘tau’, but originally called ‘heavy lepton neutrino’. This is the neutrino which, as envisaged in the original proposal, will be the result of the ‘oscillation’ from the neutrinos beamed underground from CERN and searched for at Gran Sasso with the OPERA detector.

The scientific programme for the Gran Sasso project was so complete when first proposed that, even now, no new item has been added. An additional detail not reported in your news story is that the best detector working at Gran Sasso is at present the Large Volume Detector set up to study cosmic neutrinos — a powerful liquid scintillator and tracking device, which has

been in full operation for more than 10 years. A detailed study of the time correlation and of the neutrino beam structure indicates that the neutrino beam works exactly as designed.

“Without memory there is no civilization and no physics,” as Enrico Fermi used to say.

Antonino Zichichi

National Institute for Nuclear Physics, Via Enrico Fermi, 40-00044 Frascati, Rome, Italy

Biodiversity definitions vary within the discipline

SIR — R. M. Ewers and A. S. L. Rodrigues, in Correspondence (“Speaking different languages on biodiversity” *Nature* **443**, 506; 2006), raised the issue of vocabulary differences between conservation biologists and economists, a problem that has long been recognized by biodiversity scientists.

We carried out an informal email survey of different disciplines and stakeholders involved in biodiversity conservation, asking people to define terms such as biodiversity, ecosystem function and ecosystem goods and services. We received 25 responses. The definitions varied hugely, not only between ecologists and economists, but also between them and policy-makers. Most alarming is that the definitions varied among the 13 ecologists who responded. Although this is a small sample, they were all high-profile scientists and the results were surprising.

When defining biodiversity, for example, seven used that of the Convention on Biological Diversity, two used species richness and four used other definitions.

To create solutions for biodiversity loss, it is essential for natural and social scientists to overcome such language barriers. But for this to be of any use, scientists also need to become much better at communicating their findings to policy-makers, and understanding from them the science knowledge that is required to make policy.

Alison Holt

bioSUSTAINABILITY International Project Office, University of York, Heslington, York YO10 5DD, UK

US scorn for treaties hasn't improved nuclear security

SIR — Your Editorial “A global folly” (*Nature* **443**, 605; 2006) states that scientists working on disarmament and anti-proliferation issues “have looked on aghast as disarmament sceptics around the world have scorned the value of international treaties”. To my knowledge, there was initially only one country scornfully spreading scepticism on, among others, the Comprehensive Nuclear Test Ban Treaty and the Non-Proliferation Treaty. That was the United States of America.

Sebastian Raupach

Institute of Atmospheric Physics, Johannes Gutenberg University, 55099 Mainz, Germany

BOOKS & ARTS

One culture?

We can gain a clearer picture of visual representation by crossing the divide between art and science.

Seen | Unseen: Art, Science and Intuition from Leonardo to the Hubble Telescope

by Martin Kemp

Oxford University Press: 2006. 368 pp.

£25, \$45

Bart Kahr

When British newspaper *The Independent* published a false-coloured image of Jupiter's atmosphere compiled from data taken by the spacecraft Galileo, the accompanying article began: "It may look like a Turner painting...". Martin Kemp enjoys both the absurd and the befitting in this characterization. The image of Jupiter covered 143 million square miles but was coloured in such a way as to resemble landscapes on planet Earth that we have been conditioned to recognize by Turner, who, striving to capture the "awesome infinities" of his Romantic age with scale-invariant vistas, would have exalted in seeing his Earth-bound views of nature reaching into the heavens. Why give Jupiter over to Turner's palette? In so doing, to what extent does Turner then influence how we understand and experience other worlds? And what can we learn about periods and cultures when methods of representation are set within the "broad history of visual things". These are some of Kemp's preoccupations in *Seen | Unseen*, a personal exploration of the dialogue, spoken and unspoken, between scientific and artistic representations of nature.

I began with an illustrative example of Kemp's interests as a guide for potential readers uncertain how to approach a book whose title could be about anything or everything. Authors often use the subtitle to ground tantalizingly vague titles by limiting the area of enquiry. In this case, the subtitle only makes things worse: *Art, Science and Intuition from Leonardo to the Hubble Telescope*. These are big topics for a comparatively slender volume. *Nature* readers will already be familiar with Martin Kemp, a regular author of the 'Science in Culture' columns in this journal. On the strength of his insightful essays, I would advise the reader to dive in, without worrying about the course of the river or the depth of the pool; the swim is bound to be rewarding.

Kemp is an art historian but characterizes himself as a "historian of the visual". This is, in my view, the declaration of a scholar who refuses to be told what he can and cannot look at. Throughout *Seen | Unseen*, Kemp blurs the



NASA/SPL

What are we to make of images of Jupiter rendered in the palette of painter J. M. W. Turner?

distinctions between those who have been labelled either as 'scientist' or 'artist'. He succeeds by focusing on people who defy classification. Among these are the expected giants, such as Leonardo, but also some lesser-known individuals of remarkable accomplishment, such as Bernard Palissy, a French potter obsessed with casting snakes and other creatures into his ceramics. His strange methods gave him a unique perspective on the process of fossilization, and he became a forefather of modern palaeontology. These crossover artists, and legions of others scrutinized to varying degrees, are much admired by Kemp, himself a crossover artist.

Seen | Unseen left me exhilarated at times, but perplexed at others. I found my troubles anticipated exactly by Kemp's warning in the preface: "I am aware of some problematic shifts of gear, as detailed scrutiny of a small segment

of territory at a steady pace is interspersed with generalizations that rush breathlessly across sketchily portrayed landscapes... The best I can hope is that the topics subjected to detailed focus have been selected in such a way as to be exemplary, so that the parts can be seen as belonging to a discernible whole, without using what would normally be seen as a consistent historical 'argument'." If ever there was an honest man...

So why not struggle to remove those anticipated problematic aspects? Because, in my view, this is a mature book written by a scholar who has meditated for decades on the false dichotomy between scientific and artistic representations of nature, having grown tired of superficial analyses of scientific artists and artistic scientists, and accustomed to the presentation of comparable material in the form of tightly focused magazine columns. The loose

construction of *Seen* | *Unseen* must have been a liberating exploration. If Kemp asks a lot of his readers — one should read the book straight through, if possible, to appreciate the dialogues between chapters and between characters separated by centuries — we can't begrudge him a little confusion because he invited everyone to the party he has arranged with his favourite students of nature.

"My final plea," Kemp offers, "is for profes-

sionals in each 'field' to exercise generosity and tolerance... I believe that we surrender our human 'wholes' to our specialist 'parts' at our peril." Not only is this plea to academics self-evidently sensible and bound to be rewarding if heeded, it is hard today not to read it as wisdom for the wider world. ■

Bart Kahr is in the Department of Chemistry, University of Washington, Seattle, Washington 98195, USA.

EXHIBITION

Sowing the seeds

Gregor Mendel: Planting the Seeds of Genetics

The Field Museum, Chicago, until 1 April 2007, then touring the US until September 2008.

www.fieldmuseum.org/mendel

Simon Mawer

Until now you had to travel to the Mendel Museum in Brno, Czech Republic, to see the relics of Gregor Mendel's life. But in September, the Field Museum in Chicago opened its doors to the exhibition *Gregor Mendel: Planting the Seeds of Genetics*. Visitors can witness at first hand the microscope Mendel used to make his ground-breaking observations — he was the first to demonstrate conclusively that only a single pollen grain is needed for fertilization — and the telescope with which he observed sunspots. Perhaps most poignant of all, they can see the spectacles through which the short-sighted friar viewed his world. Mendel has often been billed as something of a mystery man in the history of science, so, with genetics set to be the focus of much attention in the twenty-first century, it is appropriate that he should be the subject of a popular exhibition.

The exhibition isn't vast — unlike Darwin, Mendel left little behind — but it is beautifully designed and strikes an excellent balance between the more serious documentary side of the story and the kind of interactive display that is obligatory these days. Children can cross-breed virtual peas on screen, and, having followed Mendel's algebraic argument, open an oversized plastic pea pod to discover white and green peas in a perfect three-to-one ratio.

Their parents, meanwhile, can get to grips with lucid and informative captions that set Mendel's work in historical context and explain some of the issues surrounding his discoveries. For example, among a selection of books from the abbey library, there is Mendel's own copy of the German edition of Darwin's *On the Origin of Species*, open at a page where, tantalizingly, Mendel has marked a piece of text with pencil lines in the margin and a single, enigmatic exclamation point. This is where Darwin writes: "There are many laws regulating variation,

some few of which can be dimly seen." The irony is, of course, that at the very time that Darwin was writing those words, Mendel was shedding a dazzling light on those dimly seen laws. Had Darwin only read, and understood, Mendel's work, he would have found there the mechanism of inheritance that his theory of natural selection so desperately needed.

Many of the personal documents on display are facsimiles (the originals were considered too delicate to travel), but they are beautifully exact copies, and Mendel's meticulous meteorological records are original. So too is the letter from Carl von Nägeli, the botany professor with whom he corresponded and who famously failed to grasp the significance of the work done by the self-effacing friar from Brunn (as Brno was known during the Austrian empire). Appropriately, the Mendel section closes with a life-size portrait of Mendel as abbot, a promotion that effectively put an end to his research.

From Mendel we move on, with a nod in the direction of Francis Galton, to a brief consideration of genetics in the twentieth century. The Field Museum is a natural-history institution, so it is no surprise that this later part of the display has a deliberate natural-history emphasis, rather than the biomedical one that is now so often the stuff of genetics.

A further facet of this small but intriguing exhibition is a group of genetically inspired artworks. These range from a large photograph by Susan Degas of a honeycomb (Mendel bred bees), which looks uncannily like a DNA microarray, to Christine Borland's pedigree of a real family with Huntington's disease, transformed into a stunning mobile with agate disks to represent the individuals.

There is, however, a ghost at this feast of science, history and art, an item that is conspicuous by its absence: Mendel's famous 1865 paper that started it all, '*Versuche über Pflanzen-Hybriden*'. The exhibition has an original print version on display, but what about the manuscript copy in Mendel's own immaculate copperplate? There is a facsimile of a single sheet, but where is the real thing? The answer is, I'm afraid, we don't really know. This seminal paper, the legal property of the Abbey of St Thomas in Brno, is not in the abbey's possession. Neither is it accessible to the public. Following its disappearance from a Moravian bank vault during the Second World War (when it was photographed), the manuscript resurfaced in the 1990s in the possession of a member of the Mendel family in Germany. So the story goes. Quite why there should be such secrecy over this valuable artefact I do not know.

I suspect it is something to do with politics. ■ Simon Mawer lives in Rome and is the author of *Mendel's Dwarf* and other novels, as well as the companion book to this exhibition, *Gregor Mendel: Planting the Seeds of Genetics* (Abrams, 2006).

Corrections

In his review of *Does Measurement Measure Up? How Numbers Reveal and Conceal the Truth* by John M. Henshaw (*Nature* **442**, 357, 2006), David Colquhoun criticized the methods of ranking the quality of teaching at UK universities. In fact, the Quality Assurance Agency for Higher Education (QAA) stopped its assessment of the teaching quality of individual subjects, resulting in a graded numerical profile, in 2001, and discourages others from using its judgements in league tables.

In a review of *Reconceiving the Gene: Seymour Benzer's Adventures in Phage Genetics* by Frederick Lawrence Holmes (*Nature* **443**, 509, 2006), we omitted to mention that the book was completed and edited by William C. Summers. We apologize to Professor Summers for this oversight.



Mendel's microscope can now be seen in Chicago.

S. BARTOS

EXHIBITION

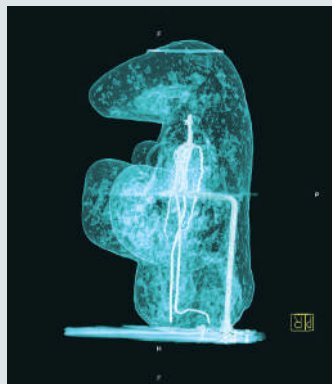
The ballerina within

At just over 60 centimetres tall, Picasso's *Bust of a Woman* (1931) is a small, compact mass of a sculpture that smacks of Stone Age totems. Using computerized tomography (CT), photographic artist Xavier Lucchesi has assembled three-dimensional images of it (shown here), revealing what appears to be another, antithetical form hidden inside. At the sculpture's heart lies a finely worked metal structure that bears a strong resemblance to a ballerina. Standing on her points, with head erect and arms



gracefully bent, she even appears to be wearing a tutu.

The tutu, it turns out, is an artefact of the technology: a halo of light similar to the phantom thrown up by a dental filling in a CT head scan. Picasso would build his sculptures



around metal skeletons, which, in other works scanned by Lucchesi, seem to have served a purely functional role. So it is possible that the ballerina wasn't Picasso's creation at all — that she exists only in the 'eye' of the scanner. If so,



there are at least two artists at work here, and one of them is a machine.

Xavier Lucchesi's exhibition, 'Picasso X-RAYS', can be seen at the National Picasso Museum in Paris (www.musee-picasso.fr) until 8 January 2007. **Laura Spinney**

SUCCESSION PICASSO/X. LUCCHESI/EDITIONS MBE

Easing the pain

The Worst of Evils: The Fight Against Pain

by Thomas Dormandy

Yale University Press: 2006. 560 pp. \$35, £19.99

John Carmody

Some years ago Australian poet Les Murray wrote a set of sonnets about his student days at Sydney University. The lines

*With Duncan the Sydney historian who
in an Australian course might send off the First Fleet
by August*

referred to Duncan McCallum, who was legendary for sometimes completing his lectures with the founding fleet still in Portsmouth. I felt that way about *The Worst of Evils* by Thomas Dormandy. Certainly human history is long, but our knowledge of the mechanisms of pain, and our capacity to treat it reasonably well, are quite recent, so it seems idiosyncratic for the author to take about 400 pages to reach the twentieth century.

In this odyssey, Dormandy tells an almost Homeric or Chaucerian saga, and if sometimes his divagations are rather too numerous, they are mostly interesting. Clearly, he is a learned man, but good — and cogent — writing demands selectiveness. You cannot put everything you know into a book; it fatigues the reader. His footnotes are numerous and sometimes florid, but even so I was often frustrated that some of his most interesting or dogmatic assertions were delivered *ex cathedra*, without substantiation. Regrettably, there are sufficient errors of fact to leave me uneasy, even when the author is at his most confident.

He is quite wrong, for example, in his account of the anatomy of the spinal cord and its nerves; he seems confused about what Otto

Loewi really did in 1921 when he discovered chemical neurotransmission; his grasp of modern neuroscience seems flawed, or perhaps just infelicitously expressed; and when his poetic muse usurps the historian's gravitas, the results are often unconvincing: "Soon the patella hammer would become as numinous a repository of medical wisdom as the stethoscope."

If, as Hippocrates observed and other writers have echoed, life is short and the art is long, the huge span of Dormandy's chronicle is hardly surprising. Nor is the fact that it seems to have as many characters as the phone book. As in a great novel, these people are vain, diligent, amorous, honourable, insightful, perverse, obtuse and arrogant. But unlike a novel, this book of 50 chapters (plus introduction and epilogue) does not have to be read all at once. As at a *yum cha* banquet, one can (and should) be choosy about the dishes. The pleasure of the book is as much in the travelling as the arriving; indeed, Dormandy seems to relish providing distractions for the reader.

The chapters on ether anaesthesia in the United States are especially enthralling. What the London surgeon Robert Liston referred to as a "Yankee dodge" was a marvellous clinical advance. But it had an unorthodox history, which Dormandy tells excellently, with the six principal characters playing out an almost Shakespearean drama. Whatever the avarice and personal tragedy involved, what apparently began as a circus act by 'Professor' Sam Colt (the inventor of the famous 'six-shooter') led to the safe anaesthesia we take for granted today, and makes possible the extraordinary range of modern surgery — even though it is the surgeons who take the glory.

Dormandy tells a good story, too, about the remarkable John Snow, who is best known for

his painstaking epidemiological study of cholera in London — he identified the source as the Broad Street pump — but is equally deserving of renown as a pioneer anaesthetist. Dormandy also gives a fascinating account of the development of phenacetin and aspirin, and points to the moral ambiguity of the liaisons between industry and what is, sometimes disingenuously, called 'disinterested research'. He writes well, too, on the first era of cocaine abuse (which carried severe risks to experimenter and patients alike) as an example of the hazards of early clinical pharmacology. And much of what he says about self-medication using proprietary 'remedies' of a bewildering diversity is fascinating social history.

Part of the challenge facing Dormandy — and, indeed, every author on this topic — is that, like the perception of beauty, pain is a personal and 'internal' matter. He quotes Humphry Davy, who in 1800 wrote about his experiments with nitrous oxide: "From the nature of the language of feeling, my description must remain imperfect... We are at the best of times incapable of describing pleasure and pain except by means of inadequate terms which have been associated with them at the moment of experiencing them." But we can say when a perception, irrespective of whether it is pleasant or noxious, has disappeared or become significantly less intense.

Clinically effective anaesthesia, then, developed without any real understanding of what pain is, or how its perception is generated. This is why the climax of Dormandy's account of that 'Yankee dodge' is so splendid, with the ringing declaration — and it still rings! — by the previously sceptical Boston surgeon John Collins Warren, after he successfully operated on an etherized patient: "Gentlemen! This is no humbug."

John Carmody is at the School of the History and Philosophy of Science, University of New South Wales, Sydney, NSW 2052, Australia.

CELL BIOLOGY

A clean energy programme

Toren Finkel

Mitochondria supply cells with energy, but in the process produce potentially damaging oxidants. It seems that a protein required to produce new mitochondria also protects against the resulting oxidative damage.

Climate scientists have taught us that the planet is getting warmer because of the intimate relationship between the production and consumption of energy and the consequent generation of toxic by-products. Cell biologists have long known that a similar struggle between product and by-product occurs continuously inside every cell. Ground zero for this intracellular battle is located in the mitochondria, tiny and evolutionarily ancient energy-producing organelles. Writing in *Cell*, Spiegelman and colleagues¹ suggest that these structures might be able to teach modern cell biologists — and those concerned about global warming — a few lessons on handling the delicate balance between producing what we need, while not ruining what we have.

Much like any factory producing widgets, mitochondria consume carbon-based fuels. Their product is ATP, the energy currency of the cell. Nonetheless, just like factory smokestacks, mitochondria also release potentially harmful by-products into their environment. For mitochondria, these toxins come in the form of reactive oxygen species (ROS) that include superoxide and hydrogen peroxide. In turn, these oxidants can interact with other radical species or with transition metals to produce by-products that are even more damaging. To combat ROS production, the cell has evolved a number of sophisticated antioxidant defences, including enzymes such as superoxide dismutase to scavenge superoxide, as well as catalase and glutathione peroxidase to degrade hydrogen peroxide.

Spiegelman and colleagues¹ show that treating cells with extra (exogenous) hydrogen peroxide stimulates the production of several antioxidant enzymes, including catalase, glutathione peroxidase and various forms of superoxide dismutase. This effect occurs

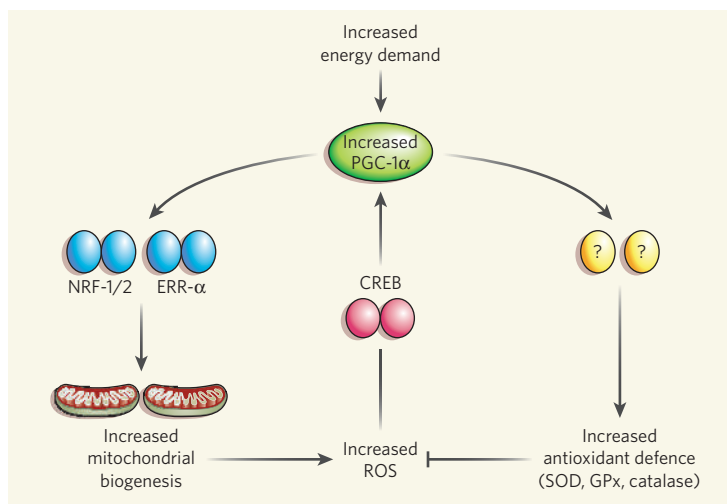


Figure 1 | PGC-1α in mitochondrial biogenesis and antioxidant defence¹⁻³. Many external stimuli that signal an increase in the need for energy can activate PGC-1α. By acting on members of the nuclear-receptor superfamily — including NRF-1, NRF-2 and ERR-α — PGC-1α can induce mitochondrial biogenesis. The resulting production of reactive oxygen species (ROS; oxidative stress) may feed back through the CREB gene-regulatory factor, leading to further increases in PGC-1α. This oxidant-induced increase in PGC-1α is then necessary to stimulate expression of a host of antioxidant proteins, including forms of superoxide dismutase (SOD), glutathione peroxidase (GPx) and catalase.

through increased transcription, the initial step of gene expression, and it seems to require a nuclear protein called PGC-1α. Spiegelman and colleagues first described PGC-1α several years ago, and it is now clear that this protein belongs to a family of coactivators that includes PGC-1β and PRC (ref. 2). Coactivators are proteins that regulate gene transcription, usually in a tissue-specific fashion. Curiously, these factors don't actually bind to regulatory sequences in DNA. Instead, they often act as molecular scaffolds to turbo-charge gene expression. They do this by recruiting various DNA-modifying enzymes while simultaneously interacting with classical transcription factors that bind to DNA. Interest in PGC-1α escalated when it was realized that this particular coactivator has a major role in creating new mitochondria when they are needed by the cell, in a process known as mitochondrial biogenesis³.

In their latest experiments¹, Spiegelman and colleagues show that simultaneously reducing the amounts of PGC-1α and PGC-1β

essentially eliminates the cell's ability to increase its antioxidant defences in response to exogenous ROS. Their experiments suggest that although PGC-1β is involved in this response, most of the transcriptional drudgery is done by PGC-1α. Indeed, they show that cells from PGC-1α-deficient animals have increased basal levels of ROS and a reduced ability to withstand damage from these toxins. Similarly, overproduction of PGC-1α in neuronal progenitor cells allowed those cells to become more resistant to oxidative stress.

These *in vitro* results were mirrored by *in vivo* observations. For instance, treatment of mice lacking PGC-1α with two known inducers of neurological oxidative stress, MPTP (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine) and kainic acid, caused significantly more neuronal damage than did similar treatment of control mice. In the 1980s, MPTP caused severe Parkinson's disease in a group of drug abusers who accidentally ingested this toxin. It is now known that in these unfortunate individuals MPTP was acting as a highly specific inhibitor of mitochondrial electron transport. Similarly, kainic acid treatment is widely used as a model of neuronal damage and epilepsy.

Together with previous work, these results suggest that PGC-1α can increase the number of mitochondria while also protecting the cell from subsequent mitochondrion-induced damage. In essence, it is as though PGC-1α is building the factory and starting the environmental clean-up at the same time. The observation that this protein can regulate the oxidative balance in cells might also provide insight into some unexplained traits observed in mice lacking PGC-1α. Such animals exhibit striking behavioural changes, including increased anxiety and hyperactivity^{4,5}. Although these behavioural changes may relate to structural changes seen in the brains of PGC-1α-deficient mice,

an independent group working on the genetic basis of anxiety has implicated ROS metabolism in this complex trait⁶. Are the behavioural changes in the PGC-1 α -deficient mice also due to an absence of antioxidant defences? If so, does this imply that what cell biologists call oxidative stress and what social scientists call psychological stress might ultimately share a common mechanism? Further study is necessary, but people who fear public speaking may take some comfort in the notion that the problem might not be in their head but rather, more specifically, in their mitochondria.

The work of Spiegelman and colleagues¹ brings up another interesting aspect of mitochondrial biology. It is well known that these organelles, whether they come from simple organisms or complex mammals, all leak measurable amounts of ROS. Experimental systems that simply increase the production of antioxidant proteins seem to be quite effective at reducing this leakage. If ROS synthesis is so bad, and a molecular solution so apparently straightforward, why has this 'design flaw' not been eradicated during the billions of years of evolution? There are many possible answers, but one is that the notion that ROS from the mitochondria are solely harmful could be incorrect. Indeed, substantial evidence exists that ROS generated in the cytoplasm could have vital signalling functions⁷, and this might also be true for oxidants derived from mitochondria^{8,9}. The new study¹ strengthens this possibility and suggests that a homeostatic loop exists between mitochondria and ROS and that this loop is, at least in part, orchestrated by PGC-1 α (Fig. 1).

Previous reports have suggested that certain life-extending strategies, such as calorie restriction, might work through the PGC-1 α -induced mitochondrial biogenesis programme¹⁰. Spiegelman and colleagues' study suggests that PGC-1 α could also alter our susceptibility to neurodegenerative conditions that are linked to mitochondrial dysfunction and oxidative stress, such as Parkinson's disease. Therefore, fine-tuning the activity of this resourceful coactivator might have a wide range of clinical benefits, including potentially allowing us to live longer and think more clearly. Not a bad set of objectives, especially if we are ultimately going to need to tackle really tricky problems like global warming. ■

Toren Finkel is in the Cardiology Branch, NHLBI, National Institutes of Health, 10/CRC 5-3330, 10 Center Drive, Bethesda, Maryland 20892, USA. e-mail: finkelt@nih.gov

CLIMATE CHANGE

The south-north connection

Eric J. Steig

A new ice-core record from Antarctica provides the best evidence yet of a link between climate in the northern and southern polar regions that operates through changes in ocean circulation.

Over the past 20 years, the analysis of ice cores has been transforming our understanding of past climate. Most notably, the Vostok core from Antarctica¹ provided remarkable evidence of the correspondence between temperature and atmospheric carbon dioxide concentrations over the past 420,000 years. And the GISP2, GRIP and NGRIP cores from Greenland^{2,3} offered a view in unprecedented detail of climate change over the past 100,000 years (including the revelation that abrupt warming events of 10 °C or more have taken place in Greenland).

More recently, the European Project for Ice Coring in Antarctica (EPICA) obtained the longest ice-core record yet⁴, one spanning 800,000 years of climate history, from Dome C, in the same sector of Antarctica as Vostok. The EPICA team has now pulled off another feat. As it reports on page 195 of this issue⁵, it has completed analysis of 2,500 metres of an ice core from Dronning Maud Land in the Atlantic sector of Antarctica (Fig. 1).

The significance of this 'EDML' core is not its sheer length; at the site concerned, 2,500 m takes us back 150,000 years. Rather, it is its resolution: this is the highest-resolution record obtained outside Greenland that extends well beyond the last glacial maximum (about 20,000 years ago). In consequence, the EPICA team has been able to place the EDML record with high precision on the same timescale as the records from Greenland. This allows us to compare the Greenland and Antarctic records over time intervals as short as a few centuries.

At this point, an analogy may help. On a year-to-year basis, much global climate variability is dominated by the El Niño–Southern Oscillation (ENSO). Understanding ENSO has required comparison of climate records in the main ENSO region (the tropical Pacific) with records elsewhere. This could not have been achieved without placing different time-series data on the same timescale. We would otherwise have little confidence, for example, in the observed correspondence between ENSO and rainfall variations in southern California. Nor would it be meaningful to try to understand the ocean and atmospheric dynamics that give rise to that relationship, because we could not reject the null hypothesis that it was merely due to chance. The Greenland and Antarctic ice-core records are likewise measures of climate variability, but on timescales of centuries, millennia and longer. Indeed, the high-resolution measures of climate afforded by ice-core records show unambiguously that climate varies on these longer timescales much more widely than one would expect from simple extrapolation of the power spectrum of observed (modern) climate⁶.

The usual explanation for the millennial-scale variability is that it is due to changes in the deep meridional overturning circulation (MOC) in the Atlantic Ocean. Put simply, a vigorous MOC is thought to deliver heat to the North Atlantic at the expense of the Southern Ocean. Increases and decreases in MOC strength should thus result in a climate 'see-saw' between the Southern and Northern

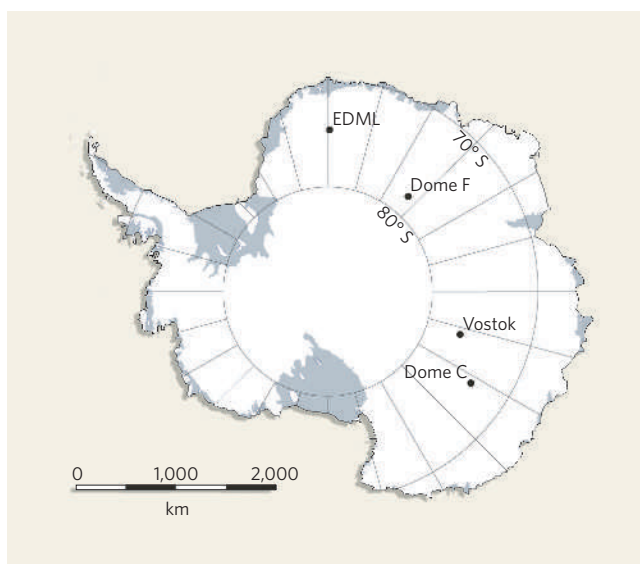


Figure 1 | Core sites. Locations of deep ice-core drilling projects in Antarctica where records longer than 100,000 years have been obtained. The new EDML core⁵ in Dronning Maud Land has the highest time resolution of these cores.

1. St-Pierre, J. *et al. Cell* **127**, 397–408 (2006).
2. Finck, B. N. & Kelly, D. P. *J. Clin. Invest.* **116**, 615–622 (2006).
3. Puigserver, P. *et al. Cell* **92**, 829–839 (1998).
4. Lin, J. *et al. Cell* **119**, 121–135 (2004).
5. Leone, T. C. *et al. PLoS Biol.* **3**, e101 (2005).
6. Hovatta, I. *et al. Nature* **438**, 662–666 (2005).
7. Finkel, T. *Curr. Opin. Cell Biol.* **15**, 247–254 (2003).
8. Nemoto, S., Takeda, K., Yu, Z. X., Ferrans, V. J. & Finkel, T. *Mol. Cell Biol.* **20**, 7311–7318 (2000).
9. Werner, E. & Werb, Z. *J. Cell Biol.* **158**, 357–368 (2002).
10. Lopez-Lluch, G. *et al. Proc. Natl Acad. Sci. USA* **103**, 1768–1773 (2006).

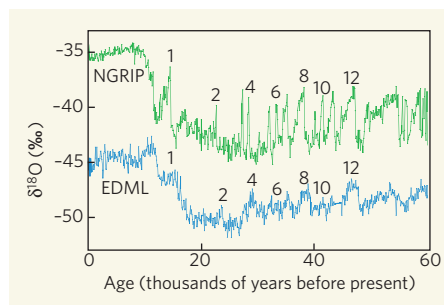


Figure 2 | Core data. Comparison of records from the EDML ice core from Antarctica⁵ and the NGRIP core from Greenland³; oxygen-isotope ($\delta^{18}\text{O}$) ratios are a measure of temperature. Temperatures increase gradually in Antarctica, reaching a maximum at about the same time that Greenland warms rapidly. The numbers refer to designated warming events in Greenland and the corresponding Antarctic temperature maxima. Statistically, the records are largely coherent for time periods down to around 800 years, which is similar to the relative dating uncertainty. Thus, the Greenland and Antarctic climates are meaningfully related on timescales as short as measurement precision allows.

Hemispheres⁷. An observation cited in support of this idea is that there is an out-of-phase relationship between Antarctic and Greenland ice-core records of temperature (or rather, of oxygen and deuterium isotope ratios, which are well-established proxies for temperature)⁸. The problem has been that — unlike in the ENSO analogy — there is considerable uncertainty in the dating. Analyses of the existing records have generally shown that the relationship between the Greenland and Antarctic records is weak and not statistically significant, except on the very longest timescales associated with well-understood astronomical factors (the Milankovich forcing of ice ages)⁹. So it has been difficult to rule out the null hypothesis that the variability in these records largely reflects regional phenomena such as variations in wind patterns or sea-ice extent.

This is where the EDML results come in. They show that the Antarctic and Greenland ice-core records are meaningfully related, and on quite short timescales. In particular, comparison of the oxygen-isotope records shows that one can make a direct link between the distinctive temperature maxima in the Antarctic record (at least going back 60,000 years) and the unambiguous abrupt warmings in Greenland (Fig. 2). Not every Antarctic temperature maximum is as distinct as the Greenland warmings; for example, it is not clear why the small maxima labelled 6 and 10 in Fig. 2 should count as 'events', but the similarly sized bumps between maxima 1 and 2 should not. But the relationship is too strong to be due to chance.

In fact, for the interval 20,000 to 90,000 years ago a remarkable 40% of the variance in the Greenland records can be explained by the EDML time series. A more rigorous estimate of the spectral coherence between the records shows that this significant relationship extends

to timescales as short as a few centuries. Furthermore, there is a consistent out-of-phase relationship between the records. They are not strictly 'antiphased', as the term see-saw would imply. Rather, the average phase relationship is about 90°. Although cold conditions in Greenland tend to be associated with warming in Antarctica, and vice versa, the peak warmth in both records actually occurs at about the same time.

Does the EDML record demonstrate the dominant influence of the MOC on climate variability? This is not just an academic question. Variations in this circulation have been invoked to explain everything from the abrupt climate changes observed in the Greenland records to the Little Ice Age (a period of cooling between about AD 1400 and 1900 in the North Atlantic region). And the possibility of a sensitive MOC has been proposed as a 'tipping point' in future human-influenced climate change¹⁰.

One objection to these ideas is that the MOC plays a minor role in the heat budget of the polar regions¹¹. Heat transport in the atmosphere is much more important, and the atmosphere might simply compensate for any changes in MOC¹². Furthermore, the causes of the purported changes in MOC are not understood. The conventional answer — flooding of the North Atlantic Ocean by ice and meltwater from the Laurentide ice sheet in northern North America (so-called Heinrich events) — is not very convincingly supported by the evidence¹³.

The EDML data do not directly address these concerns. But they are nonetheless compatible with the idea that the MOC has a central role in millennial-scale variability. What is particularly compelling is that there is a strong linear relationship between the magnitude of warming in Antarctica and the

duration of the warm period that follows each abrupt event in Greenland (see Fig. 3 of the paper⁵ on page 197). The authors' explanation is simple: the duration of the warm periods in Greenland reflects the duration of reduced MOC, and hence the amount of heat retained in the Southern Ocean. This is consistent with a model¹⁴, proposed a few years ago, in which the magnitude of Antarctic temperature change is controlled by the effective size of the Southern Ocean heat reservoir (including both dynamic and thermodynamic effects). We may have to wait some time before we see whether these results can be reproduced by more sophisticated ocean-atmosphere climate models, because realistically encapsulating the dynamics of the Southern Ocean in such models remains a problem. But we can hope that these new results⁵ will inspire the relevant work to be done.

Eric J. Steig is in the Department of Earth and Space Sciences, 70 Johnson Hall, University of Washington, Seattle, Washington 98195, USA. e-mail: steig@ess.washington.edu

1. Barnola, J. M., Raynaud, D., Korotkevich, Y. S. & Lorius, C. *Nature* **329**, 408–414 (1987).
2. Grootes, P. M., Stuiver, M., White, J. W. C., Johnsen, S. & Jouzel, J. *Nature* **366**, 552–554 (1993).
3. North Greenland Ice Core Project members *Nature* **431**, 147–151 (2004).
4. EPICA Community Members *Nature* **429**, 623–628 (2004).
5. EPICA Community Members *Nature* **444**, 195–198 (2006).
6. Huybers, P. & Curry, W. *Nature* **441**, 329–332 (2006).
7. Crowley, T. J. *Paleoceanography* **7**, 489–497 (1992).
8. Blunier, T. et al. *Nature* **394**, 739–743 (1998).
9. Roe, G. H. & Steig, E. J. *J. Clim.* **17**, 1929–1944 (2004).
10. Rahmstorf, S. in *Encyclopedia of Quaternary Sciences* (ed. Elias, S. A.) (Elsevier, Amsterdam, in the press).
11. Seager, R. & Battisti, D. S. in *Global Circulation of the Atmosphere* (eds Schneider, T. & Sobel, A. H.) (Princeton Univ. Press, in the press).
12. Wunsch, C. *Quat. Res.* **65**, 191–203 (2006).
13. Steig, E. *Nature* **439**, 660 (2006).
14. Stocker, T. F. & Johnsen, S. J. *Paleoceanography* **18**, 1087 (2003).

STRUCTURAL BIOLOGY

Enzyme theory holds water

Matthew Freeman

Intramembrane proteases have attracted much attention because of their biological and medical value. The first crystal structure of one of these enzymes begins to solve the mystery of how they work.

Proteases are some of the most potent tools to which a cell has access, because, unlike most other protein-modifying enzymes, they catalyse an essentially irreversible reaction — the breaking of peptide bonds. This can obviously be used to degrade unwanted proteins, but proteases also have many vital regulatory functions¹. In the past few years, a class of proteases called intramembrane proteases has been discovered. These comprise diverse families, but all have multiple transmembrane domains and

an active site apparently embedded within the hydrophobic region of cell membranes.

All intramembrane proteases share the curious property of cutting their membrane-spanning protein targets in the transmembrane regions — that is, they are thought to perform the peptide-cleaving reaction (which requires water) within the lipid bilayer of cellular membranes². As this is an environment where water is traditionally assumed to be in short supply, this is a somewhat heretical idea. This

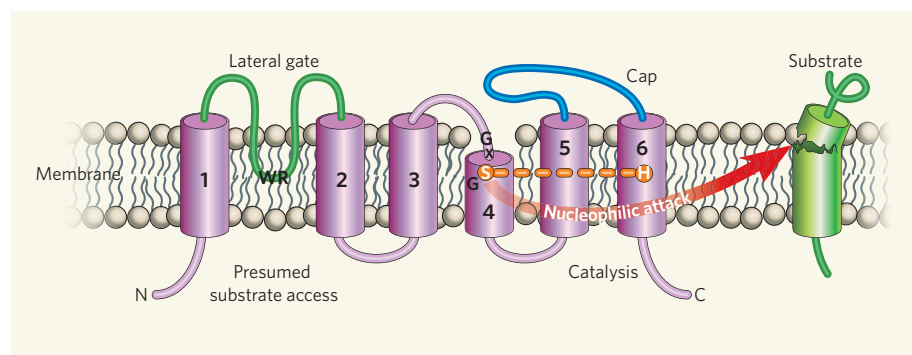


Figure 1 | Architecture of a rhomboid. The GlpG enzyme has six transmembrane helices (shown as columns)³. Helix 4 is shorter than the others, forming a depression that is open to water at the top of the membrane. The active site is at the bottom of this depression. The loop between helices 1 and 2 may form a lateral 'gate' that regulates access to the active site; the WR motif is highly evolutionarily conserved, and mutations in it at least partially inhibit enzyme activity^{4,9}. The loop between helices 5 and 6 may cap the active site, preventing access by soluble peptides. The catalytic dyad residues serine (S) and histidine (H) are joined by a hydrogen bond (dotted line); they are portrayed cleaving a hypothetical substrate by the process of nucleophilic attack. In almost all rhomboids the catalytic serine occurs in a characteristic GxSG motif^{4,6}, where G is glycine and x is any other amino acid. In GlpG the 'x' is a leucine.

has contributed to intramembrane proteases being only slowly accepted into the fold of classical proteases, despite their essential roles in various biological and medically relevant processes. In a landmark paper on page 179 of this issue, Wang *et al.*³ present the first high-resolution structure of an intramembrane protease*. At last the doubters can be satisfied, and we can begin to understand these enigmatic enzymes in atomic detail.

The rhomboid family were proposed to be serine-type intramembrane proteases⁴. Their name, and the first to be identified, comes from *Drosophila* Rhomboid-1. This enzyme cuts the membrane-tethered precursor of epidermal growth factor within the membrane, thereby releasing the extracellular domain⁴, which then signals to surrounding cells, triggering a variety of developmental decisions. Rhomboids are found in almost all organisms^{5,6} and, although little is known about their function in most species, they have been implicated in biological processes as diverse as mitochondrial function, invasion of host cells by apicomplexan parasites (including *Plasmodium* and *Toxoplasma*), bacterial communication and, as described above, growth factor signalling⁷.

So what do we learn from Wang and colleagues' structure³ of GlpG, the rhomboid protein from the bacterium *Escherichia coli* (Fig. 1)? Perhaps the main news is that it does, as predicted⁴, resemble the classical serine proteases in two key ways: it uses the amino acid serine as the means to attack the substrate peptide bond, and this serine is activated by interaction with a neighbouring histidine, which acts as a base, stripping a proton from the serine.

Beyond those features, GlpG looks rather different. The classical serine proteases rely on a famous 'catalytic triad' of amino acids, which most biochemistry undergraduates learn about

in their first year, where an aspartate has a crucial role in stabilizing the catalytic histidine⁸. By contrast, the structure implies that GlpG makes do with a serine-histidine dyad. This had been predicted from studies where the protein was mutated⁹, but it was necessarily a tentative suggestion until the structure of the active site could be examined.

Wang *et al.*³ highlight a number of features that may stabilize this dyad in the absence of the usual aspartate. First, the serine and histidine are linked by a strong hydrogen bond. Second, the histidine is also hydrogen bonded through water to another neighbouring amino acid. And third, the rotational freedom of the histidine is restricted by the aromatic ring of a nearby tyrosine. Unlike the catalytic serine and histidine, these three features are not universally conserved, suggesting that they represent secondary mechanisms that may vary between rhomboids. Overall, it seems rhomboids converged through evolution on the same activated serine catalysis as the classical serine proteases, but their machinery is stabilized in a different way.

With this structure, the problem of water — the concern that a protease should not work in a lipid bilayer — evaporates. The central transmembrane helix, which includes the catalytic serine, turns out to be shorter than predicted. This produces a hydrophilic dent at the surface of the membrane, adjacent to the active site, ensuring that there is plenty of access for water to complete the peptide-breaking reaction. Other features include a well-conserved loop between the first two transmembrane domains, which, instead of sticking out of the membrane as previously assumed, folds back within it, in a potentially unstable conformation, where it is proposed to form a substrate gating mechanism. Finally, another loop may act as a 'cap', preventing access of soluble peptides to the active site³.

Although it is tempting to believe that such a

high-resolution view of an enzyme can explain how it works, that's far from true. An X-ray crystal structure gives a static view of — in this case — an enzyme without its substrate, meaning that mechanistic conclusions involve significant inference. The old cliché that many questions remain to be answered therefore applies. How does the substrate access and bind to the active site? How is rhomboid activity regulated? Might different lipids modulate enzyme activity¹⁰, and if so, how? Can selective inhibitors be designed? And these are just for starters.

The structure of this rhomboid does not tell us much directly about other classes of intramembrane protease such as presenilin/ γ -secretase, site-2 protease, or the signal peptide peptidase family — they are unrelated to rhomboids by mechanism, sequence or evolution. It does, however, imply that intramembrane proteolysis is not a fanciful notion and, given the rather similar transmembrane architecture of all intramembrane proteases, it seems likely that they will share common strategies. Intriguingly, however, two low-resolution structures of γ -secretase have recently been published^{11,12}. Although they show rather different outline shapes of the enzyme, they agree on the existence of a central chamber, which could provide the necessary aqueous environment for catalysis. If confirmed with higher-resolution structures, this suggests that rhomboids and γ -secretase might have chanced on different structural solutions for getting water to the catalytic site.

Ultimately, what makes this rhomboid structure particularly exciting is that it gives us the first detailed view of an intramembrane protease, and that the structure correlates satisfyingly with earlier genetic and biochemical studies. A picture of how rhomboids work begins to emerge — these new kids on the block have become respectable and are well on their way to joining the establishment. ■

Matthew Freeman is at the MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK.
e-mail: mf1@mrc-lmb.cam.ac.uk

1. Ehrmann, M. & Clausen, T. *Annu. Rev. Genet.* **38**, 709–724 (2004).
2. Wolfe, M. S. & Kopan, R. *Science* **305**, 1119–1123 (2004).
3. Wang, Y., Zhang, Y. & Ha, Y. *Nature* **444**, 179–183 (2006).
4. Urban, S., Lee, J. R. & Freeman, M. *Cell* **107**, 173–182 (2001).
5. Wasserman, J. D., Urban, S. & Freeman, M. *Genes Dev.* **14**, 1651–1663 (2000).
6. Koonin, E. V. *et al. Genome Biol.* **4**, R19 (2003).
7. Freeman, M. *Nature Rev. Mol. Cell Biol.* **5**, 188–197 (2004).
8. Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (Freeman, New York, 1999).
9. Lemberg, M. K. *et al. EMBO J.* **24**, 464–472 (2005).
10. Urban, S. & Wolfe, M. S. *Proc. Natl Acad. Sci. USA* **102**, 1883–1888 (2005).
11. Lazarov, V. K. *et al. Proc. Natl Acad. Sci. USA* **103**, 6889–6894 (2006).
12. Ogura, T. *et al. Biochem. Biophys. Res. Commun.* **343**, 525–534 (2006).

*This article and the paper concerned³ were published online on 11 October 2006.

NEUROBIOLOGY

Right timing for retina repair

Thomas A. Reh

Transplants of photoreceptor cells offer hope for treating retinal disease. But getting the cells to make the right connections with the brain has been problematic. It seems the developmental stage of the cells may be the key.

Each year many people lose their sight through disease of the retina. Millions are affected by macular degeneration and retinitis pigmentosa alone. In these diseases, the cone and rod photoreceptors — the cells that convert light into neural signals — degenerate. And once these cells are lost, they are not replaced in the mammalian retina. Treatments are being developed that may delay or prevent the loss of retinal neurons in these disorders. But for those who have already lost their sight, the possibility of retinal transplantation may be the best prospect for the restoration of vision through cell-replacement therapy. Despite many attempts, such transplants have yet to produce better vision in mammals because the transplanted cells do not wire up to the brain properly. In this issue, MacLaren *et al.* (page 203)¹ show in mice that the trick may be to use cells at a particular stage in their development.

Retinal transplantation has a long history. Classic studies in the 1920s showed that transplantation of the eyes of normal salamanders could give vision to blind, cave-dwelling salamanders². This work allowed developmental biologists to explore the mechanisms that enable the correct connections to form between

the eyes and the nervous system. The first successful transplant of a mammalian retina dates back to 1959, when Royo and Quay³ transplanted fetal rat retinas into the eyes of adults of the same strain. Although the transplanted retinas did not seem to connect with the host retinas, they survived for months.

Since then, intact retinal sheets from embryonic mice and rats have been transplanted to the sub-retinal space. There they develop many characteristics of a normal retina^{4–6}, and grow to form a second retinal layer underneath the host retina. Also, transplants of 'microaggregates' — clumps of a few retinal neurons — from newborn mice developed most characteristics of rod photoreceptors, including the expression of the rhodopsin protein and even the characteristic structures called outer segments^{7,8}. Unfortunately, both the intact embryonic retinal sheets and the microaggregates of photoreceptors keep to themselves, without interacting or integrating very effectively with the host retinal neurons.

Integration with the host retina is much better when transplanting retinal progenitor cells — the immature cells that are responsible for producing all the retinal cells during embryonic development. These progenitors (also

called retinal stem cells by some) are derived from fetal or newborn mice or rats, or from human fetuses. They can be maintained in cell culture and continue to proliferate and generate new neurons and specialized retinal support cells called Müller glia^{9–11}. When these cells are transplanted into either normal or degenerated (dystrophic) retinas of rats and mice, they can migrate into all retinal layers and develop morphological characteristics of various retinal cell types^{12–15}. However, the cells do not seem to integrate efficiently into the outer nuclear layer, where the rods and cones reside, and, for the most part, evidence showing expression of photoreceptor-specific genes in the transplanted cells has been lacking (but see ref. 15).

So, MacLaren *et al.*¹ aimed to combine the migration and integration potential of the progenitor cells with the photoreceptor differentiation properties of the retinal sheets. In the first set of experiments, they tested whether retinal progenitor cells were in fact better at migrating into the host retina than mature, differentiated neurons. Using cells taken from the retina of embryonic mice or newborn mice at different postnatal ages, they found, surprisingly, that the cells that integrated into the outer nuclear layer most effectively were the postnatal cells. This was despite the fact that the embryonic cells included the highest percentage of progenitor cells.

The transplanted rod cells developed morphology identical to normal host rods, even making clear outer segments. On further analysis, the authors found that the best ages for the donor cells were from postnatal day 3 to postnatal day 5, roughly corresponding to two or three days following the peak of

PHYSIOLOGY

Channelled pain

It seems an unlikely connection. But as they report in this issue, David Julius and colleagues have identified a common factor in the way that a tarantula spider and a chilli plant defend themselves (J. Siemens *et al.* *Nature* **444**, 208–212; 2006).

The particular tarantula is *Psalmopoeus cambridgei* (pictured), a native of the West Indies. From investigations of the spider's venom, Julius and colleagues have identified new members of the 'inhibitor cysteine knot', or ICK, family of peptide toxins.

The newly isolated peptides are dubbed 'vanillotoxins' because of their action on the TRPV1 vanilloid receptor, and they cause pain and inflammation when injected into mice. TRPV1 is a member of the TRP group of cell-membrane ion

channels and was first identified through its sensitivity to capsaicin, the red-hot component of chilli peppers. Unlike almost every other ICK peptide, however, vanillotoxins activate rather than inhibit their target — in this case, TRPV1 channels in sensory neurons.

An obvious question is whether the excitatory effect on TRPV1 means that the vanillotoxins have a different mechanism of action from the other ICK peptides. A comparison with ICK peptides that have similar amino-acid sequences — hanatoxins and heteroscodratoxins, which inhibit voltage-gated potassium channels — suggests that they don't. For instance, two of the vanillotoxins can also inhibit voltage-gated potassium channels; like hanatoxins and heteroscodratoxins, they impede the



channel mechanism so that larger changes in voltage are required for the channel to open. Parenthetically, then, these data lend further support

to the idea that TRP channels and voltage-gated potassium channels are structurally related.

Lesley Anson

rod photoreceptor production in the mouse retina. This suggested that, contrary to expectations, newly born rods, rather than the progenitor cells, were the best for transplanting. The authors further corroborated this idea by harvesting just newly born rods immediately after they are generated by the progenitors and using them for transplants.

In addition to the striking morphological evidence for integration of the transplanted rod photoreceptors into the normal outer nuclear layer, MacLaren *et al.*¹ provide evidence that the transplanted photoreceptors connect up with the host retinal neurons, and can partly restore the response to low light levels in mice that are blind because they are deficient in the rod protein rhodopsin.

These results provide the best evidence so far that cell-replacement therapy may be possible. But there's a catch. If this scenario were to be applied to humans, one would have to obtain newly generated rods from the stage of development comparable to post-natal days 3–7 in the mouse. This is likely to be in the second trimester and is clearly not

feasible. However, a recent report from our lab shows that cells expressing the protein Nrl — a marker of newly born rods — can be obtained from human embryonic stem-cell lines, given certain culture conditions¹⁶. Further studies will of course be needed to determine whether the stem-cell-derived Nrl⁺ cells behave like the Nrl⁺ mouse rod photoreceptors before these can be considered for cell-replacement therapy.

Although these results provide hope for the cell-based treatments of retinal disease, they also have implications for transplantation strategies in other areas of the central nervous system. Transplantation of neural tissue is a potential treatment for several degenerative neurological diseases and traumatic injury. The studies attempting this approach used various ages of developing neural progenitors and stem cells. But it may be that the specific time at which the particular cell is harvested will make all the difference in its potential for integration and functional differentiation following transplantation. Sometimes, timing is everything. ■

Thomas A. Reh is in the Department of Biological Structure, 357420 Health Sciences Center, University of Washington, School of Medicine, Seattle, Washington 98195, USA.
e-mail: tomreh@u.washington.edu

1. MacLaren, R. E. *et al.* *Nature* **444**, 203–207 (2006).
2. Stone, L. S. *Invest. Ophthalmol.* **3**, 555–565 (1964).
3. Royo, P. E. & Quay, W. B. *Growth* **23**, 313–336 (1959).
4. Seiler, M., Aramant, R., Ehinger, B. & Adolph, A. R. *Exp. Eye Res.* **51**, 225–228 (1990).
5. Ghosh, F. & Ehinger, B. *Ophthalmologica* **214**, 54–69 (2000).
6. Zhang, Y., Arner, K., Ehinger, B. & Perez, M. T. *Invest. Ophthalmol. Vis. Sci.* **44**, 324–331 (2003).
7. Gouras, P., Du, J., Kjeldbye, H., Yamamoto, S. & Zack, D. J. *Invest. Ophthalmol. Vis. Sci.* **35**, 3145–3153 (1994).
8. Gouras, P. *et al.* *Invest. Ophthalmol. Vis. Sci.* **32**, 3167–3174 (1991).
9. Anchan, R. M., Reh, T. A., Angello, J., Balliet, A. & Walker, M. *Neuron* **6**, 923–936 (1991).
10. Reh, T. A. & Levine, E. M. *J. Neurobiol.* **36**, 206–220 (1998).
11. Troppe, V. *et al.* *Science* **287**, 2032–2036 (2000).
12. Chacko, D. M., Rogers, J. A., Turner, J. E. & Ahmad, I. *Biochem. Biophys. Res. Commun.* **268**, 842–846 (2000).
13. Klassen, H. J. *et al.* *Invest. Ophthalmol. Vis. Sci.* **45**, 4167–4173 (2004).
14. Coles, B. L. *et al.* *Proc. Natl Acad. Sci. USA* **101**, 15772–15777 (2004).
15. Qiu, G. *et al.* *Exp. Eye Res.* **80**, 515–525 (2005).
16. Lamba, D. A., Karl, M. O., Ware, C. B. & Reh, T. A. *Proc. Natl Acad. Sci. USA* **103**, 12769–12774 (2006).

ASTRONOMY

A troupe of near dwarfs

Sidney van den Bergh

Faint satellite galaxies of the Milky Way are being discovered that are dimmer than some of the Milky Way's star clusters. This finding poses a fundamental question: what are galaxies?

Of his 1938 discovery of two tiny satellite galaxies accompanying our own Milky Way, the astronomer Harlow Shapley wrote¹: “Presumably the gamut of galaxies had already been run. All forms had long been fully described. There were spirals, spheroidals, irregulars, with many variations on the spiral theme. The newly found organizations in [the dwarf galaxies] Sculptor and Fornax did not seem essential in order to fill in a natural sequence; they were not logically necessary. On the contrary, they introduced some doubt into the picture we had sketched — they suggested that we may be farther than we think from understanding the world of galaxies.”

Seven decades later, these words still ring true. We now know that ‘dwarf spheroidals’ of the type discovered by Shapley are in fact the most common type of galaxy in the Universe. And, in a paper to be published in *The Astrophysical Journal*, Belokurov *et al.*² report the discovery, in data from the Sloan Digital Sky Survey, of four more dwarf spheroidal satellites of the Milky Way system. One of these, the Coma dwarf galaxy, is the faintest galaxy ever observed, and is two orders of magnitude fainter than the brightest clusters of old stars known within the Milky Way. So where exactly

does one draw the line between a bright star cluster and a faint galaxy?

The obvious distinguishing criterion, size, is not always reliable: although galaxies are usually larger than star clusters, tidal forces can strip off the large envelopes of galaxies, leaving behind only a compact rump. So-called dark matter, on the other hand, does provide a reliable way of telling a galaxy from a star cluster. All galaxies are thought to be embedded in a halo of this type of matter, which emits no light and so cannot be observed directly; no dark matter has ever been detected in a star cluster.

Because they have no dark matter contributing to their overall mass, star clusters that formed from fragments of disintegrating tidal arms are expected to have small ratios of mass to light emitted. This implies that they have a lower spread of internal velocities than galaxies, as this spread is intimately connected to a body's mass. Individual stars in star clusters also all formed at roughly the same time, and therefore have similar metal abundances, whereas stars in galaxies, which require a significantly longer time to assemble, can have a large range of metallicities. But applying these criteria to faint, distant clusters is difficult, as

even the largest telescopes require a great deal of observing time to measure metallicities and internal velocities. The metallicity-dispersion criterion is also likely to be useless for the least luminous dwarfs, as these are expected to contain only very metal-poor stars.

Of the five objects discovered altogether by Belokurov *et al.*², three (dubbed Canes Venatici II, Hercules and Leo IV) have standardized radii of more than 100 parsecs (1 parsec is 3.26 light years). They are therefore large enough to be definitely regarded as faint galaxies. But the smallest of the objects, Segue 1, has a radius of only around 30 parsecs, placing it at the upper bound of the size range of known Galactic star clusters. The fifth object, the Coma dwarf, is of an intermediate size between the largest known star cluster and the smallest known galaxy. Only velocity-spread measurements will be able to tell if this object truly is a galaxy containing dark matter, or a huge star cluster that does not.

If we add the objects found by Belokurov *et al.*², and the star clusters Omega Centauri and NGC 2419, which might be the stripped cores of dwarf spheroidals, the Milky Way has 21 known companions within a distance of around 500 kiloparsecs, equivalent to about 1.6 million light years. The three most luminous of these — the Large Magellanic Cloud, the Small Magellanic Cloud and an object known as NGC 6822 — might actually be, or at least have started their lives as, independent members of the Local Group of galaxies. This group is centred on the Milky Way and M31, the great Andromeda spiral. By coincidence, the number of companions presently known within 500 kiloparsecs of Andromeda's nucleus is also 21; but some of these too might be

free-ranging members of the Local Group, rather than satellites of their central galaxy.

A curious fact, first noted by Jaan Einasto and colleagues³ in 1974, is that most of the close satellites of M31 and the Galaxy are elliptical or spheroidal, whereas many of their more distant companions have irregular forms. Unfortunately, some of the newly discovered objects² have such low star counts, and the subtraction of light from background sources is therefore so uncertain, that it is difficult to be sure whether they are irregular or dwarf spheroidal galaxies.

Another interesting facet that emerges from these latest discoveries of satellite galaxies is that the radial distances of the known companions of M31 and the Milky Way from the nucleus of their parent galaxy seem to be roughly similarly distributed. For distances of less than 150 kiloparsecs, the number N of companions within a radius R is reasonably well represented by a logarithmic correlation, $\log N = -0.24 + 0.80 \log R$. Perhaps because of the incompleteness of the data, above 150 kiloparsecs the cumulative number drops below that expected from this relation. Even so, half of the known companions of these two galaxies are situated more than 90 kiloparsecs from their host. At that distance, gravity from

an unseen source must be acting to keep the companion galaxies in tow, attesting to the enormous size of the haloes of dark matter that must surround the Milky Way and M31.

Numerical simulations^{4,5} have predicted the existence of a swarm of smaller, denser dark-mass haloes embedded within the general dark halo of the Milky Way. The gravity of each of these mini-dark-haloes should have attracted a clump of ordinary, visible matter. Thus, these predictions seemed to conflict with the observation that our Galaxy is surrounded by only a handful of companions. But the faintness of Belokurov and colleagues' new discoveries² indicates that they might be just the first of a vast population of ultra-faint dwarf galaxies surrounding the Milky Way system that is yet to be discovered.

Sidney van den Bergh is at the Herzberg Institute of Astrophysics, 5071 West Saanich Road, Victoria, British Columbia V9E 2E7, Canada. e-mail: sidney.vandenbergh@nrc-cnrc.gc.ca

1. Shapley, H. in *Galaxies* 141 (Blakiston, Philadelphia, 1943).
2. Belokurov, V. *et al. Astrophys. J.* in the press; preprint available at www.arxiv.org/astro-ph/0608448 (2006).
3. Einasto, J., Saar, E., Kaasik, A. & Chernin, A. D. *Nature* **252**, 111–113 (1974).
4. Klypin, A., Kravtsov, A. V., Valenzuela, O. & Prada, F. *Astrophys. J.* **522**, 82–92 (1999).
5. Moore, B. *et al. Astrophys. J.* **524**, L19–L22 (1999).

IMMUNOLOGY

Exposure of an executioner

Michael Carroll

The complement C3 protein binds to pathogens, singling them out for execution by the immune system. Structural studies show how the chemical group responsible for this binding is exposed on activation.

Some of the oldest molecules involved in defence in the animal kingdom are found in the thioester protein family, which dates back more than 500 million years¹. The power of these proteins comes from their thioester chemical group — a highly reactive molecular 'warhead' that, when activated, binds to chemical acceptor groups on many pathogens and marks them out for destruction by immune cells. Unfortunately, similar acceptor groups are found in host tissues too, so the thioester group must be carefully regulated to prevent it from interacting with host cells and wrongly drawing the immune system to attack them. Perhaps because of this potential for damage, the modern vertebrate immune system retains only a few thioester proteins, including complement protein C3.

Understanding how this potentially dangerous but useful defence protein is kept under control is of great interest, and X-ray crystallographers have been trying to puzzle out its structure for several decades. In September 2005, Janssen *et al.*² reported the structure of the human native C3 — the form of the

protein before activation. Now, three papers* starting on page 213 of this issue^{3–5} report the first X-ray structures of C3b, the active form of human C3.

Complement C3 is a principal component of the complement system, a large family of blood serum proteins and cell-surface receptors involved in the recognition of pathogens and directing the immune response against them, and in protecting host cells from that reaction. C3 is activated in three ways (Fig. 1, overleaf). In the 'classical pathway', it is guided to its target and activated following specific recognition of the pathogen by antibody⁶. Or, it can be triggered through the actions of lectin, a protein that binds to the sugars found on the surface of many pathogens (the 'lectin pathway')⁷. Finally, in the less specific 'alternative pathway', C3 is spontaneously activated at a low level. This results in 'painting' of tissues unprotected by the cell-surface regulators of complement⁸. The alternative pathway not only initiates C3, but also

*This article and the papers concerned^{3–5} were published online on 15 October 2006.



50 YEARS AGO

Lord Halsbury introduced the subject [of automation] by saying that the public is substantially misinformed about it... "The automatic factory and the automatic office," he said, "have arrived, but they are no more than the development of improvements in production that have been going on for many years. The only really new technique which has been injected into industry is the digital computer." He said that automation is most likely to affect the section of industry which produces what are called consumer durables, that is, motorcars, radio, television, washing machines, etc. It is very unlikely to be widely used in heavy engineering, shipbuilding, agriculture, mining, textiles and highly mechanized industries... He believes that although automation will cause men and women to change their jobs and perhaps move to new areas, widespread unemployment is not a danger because in any event changes are unlikely to come about quickly.

From *Nature* 10 November 1956.

100 YEARS AGO

Two events during the past few days have shown that men of science recognise the ability of women to originate and carry out scientific research and inspire others with their spirit. One is that on Thursday last the Royal Society awarded the Hughes medal to Mrs W. E. Ayrton, for her experimental investigations on the electric arc and also upon sand ripples; and the other event is the first lecture delivered at the Sorbonne on Monday by Mme. Curie... The Royal Society, by placing Mrs Ayrton's name alone, and not bracketed with that of a man, in the list of medallists for this year has manifested its recognition of individual work by a woman. The Davy medal was awarded by the society in 1903 to Prof. Curie and Mme. Curie jointly, for their researches on radium, though the published work on the subject shows that the discovery of radium was due to Mme. Curie alone.

From *Nature* 8 November 1906.

50 & 100 YEARS AGO

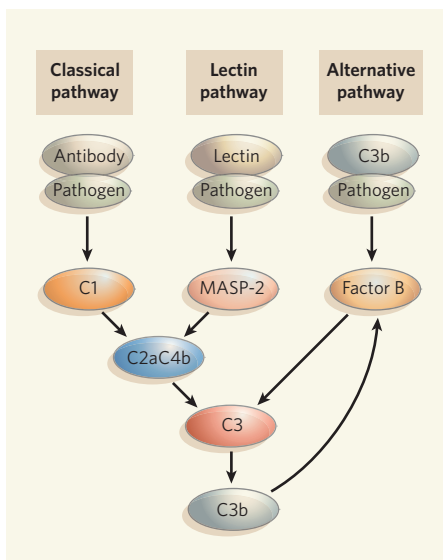


Figure 1 | Complement C3 is activated by three pathways. In the classical pathway, antibody bound to pathogen surface leads to the activation of C2aC4b. This is a 'convertase' enzyme that cleaves C3 into C3a and C3b — the active form (see also Fig. 2). Similarly, the lectin pathway is triggered following recognition of pathogens by the lectin protein, which binds to sugar molecules on their surface. This results in activation of the enzyme complex MASP2 (mannan associated serine protease 2) that forms the C2aC4b convertase. The alternative pathway not only amplifies the classical pathway and the lectin pathway but it can also be activated spontaneously. C3b bound to pathogen surfaces interacts with the blood serum protein factor B, which forms an alternative convertase called C3bBb.

amplifies the other pathways. So, understanding how C3 interacts with initiators of the alternative pathway such as factor B and regulators such as factor H would be of great value.

In the unactivated C3, the thioester warhead

is kept wrapped up out of harm's way inside the protein until it is required. In the native C3 structure², the thioester domain (TED) is tucked into the body of the protein, beneath the CUB domain (Fig. 2). The protein is triggered when it is cleaved into two fragments, C3a and C3b, and the three papers^{3–5} provide striking evidence of how the thioester is exposed at this point. The TED shifts a massive 85 Å following activation, extending out from the body of the protein to resemble a ball on a chain.

This substantial change not only exposes the thioester site, but also opens up binding sites for factor B, which leads to amplification of C3b levels through the alternative pathway (Fig. 1). Notably, sites for complement regulators such as factor H are also exposed along the stretched CUB domain. The interaction with factor H leads to cleavage of C3b within the CUB domain, shutting off further activity and resulting in fragments iC3b and C3c.

But this is not the end of C3's influence — its breakdown products have multiple functions in immunity. C3a, which is released during activation, has potent proinflammatory effects, causing constriction of smooth muscle and activation of leukocyte cells when it binds to cell-surface receptors. The activated product, C3b, is a ligand for multiple receptors including the recently identified CRIg, which helps to get rid of potential pathogens. The structure presented by Wiesmann *et al.*⁵ is of C3b in complex with CRIg, and it shows that this interaction shuts off further activity in the alternative pathway by interfering with C5 convertase (the enzyme complex of C3bBb). Notably, it does not seem to interfere with the classical pathway, presumably because of structural differences between the classical- and the alternative-pathway convertases. The authors explored the functional relevance of the C3b–CRIg interaction in a mouse model

of collagen-induced arthritis, in which the alternative pathway is a major mediator of injury. Treating the mice with a soluble form of CRIg blocked the arthritis inflammation.

The first product of C3b inactivation, iC3b, is the target of several receptors found on circulating leukocytes, stimulating these cells to engulf pathogens and to activate inflammatory cells. The final breakdown product, C3dg, targets pathogens to surface receptors on B lymphocytes, cells that are vital in ramping up the production of antibodies against the pathogen and in remembering it in case of future encounters⁹.

What happens when C3b is not effectively regulated and rapidly shut off? Lessons come from mouse models of immune deficiencies, such as a strain that lacks factor H, where the defunct regulation causes spontaneously active C3, leading to fatal kidney disease¹⁰. People bearing a genetic variant of factor H have a high susceptibility to age-related macular degeneration, a condition that leaves them blind, presumably because of their limited ability to shut off the active C3b and prevent its amplification by the alternative pathway¹¹. A variety of other inflammatory diseases have also been categorized as complement-dependent. Yet there are few generally accepted treatments available to damp this system down.

Now, with the knowledge of the structure of the active form of C3b, there is the promise of therapies to manipulate the complement system. For example, identification of the contact sites of C3b with factor B, which is crucial for initiation of the alternative pathway, should lead to effective inhibitors to block this key interaction. Also, given that the interaction of CRIg with C3b shuts off the alternative pathway, the soluble form of the receptor has potential as a treatment for disorders known to involve the alternative pathway, such as a kidney disorder called membranoproliferative glomerulonephritis type II, and possibly age-related macular degeneration.

Millions of years of evolution provided a great deal of control of the complement system, but it still occasionally runs amok. These structures of C3b could provide us with the means to bring it back in line when it does.

Michael Carroll is at The CBR Institute for Biomedical Research, Harvard Medical School, 800 Huntington Avenue, Boston, Massachusetts 02115, USA.

e-mail: carroll@cbr.med.harvard.edu

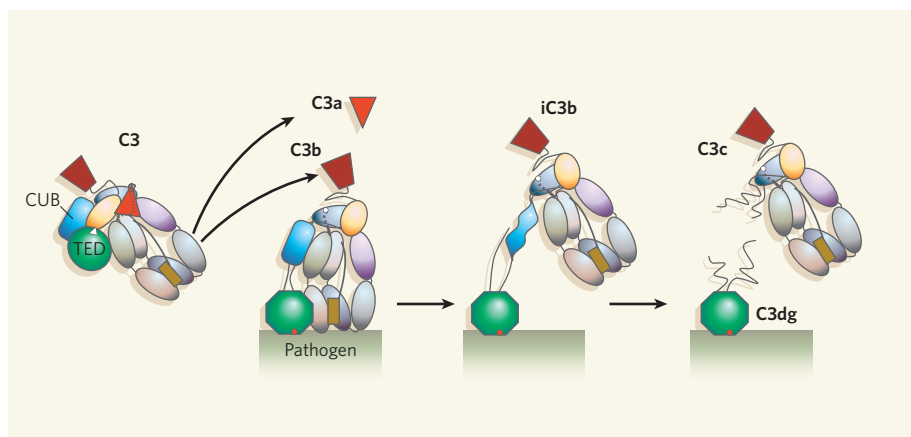


Figure 2 | Breakdown of complement C3. This model is based on the structures reported by Janssen *et al.*³, Ajees *et al.*⁴ and Wiesmann *et al.*⁵. Cleavage of native C3 protein by the enzyme C3 convertase releases C3a and induces a major shift in the overall conformation of the remaining C3b molecule. Full exposure of the thioester domain (TED) facilitates binding of the thioester (red dot) to a target pathogen surface. The opening up of C3b exposes binding sites for components of the alternative pathway such as factor B as well as regulators of complement such as factor H. These regulators aid the sequential degradation of C3b to iC3b, C3c and C3dg. C3c is inactive, but each of the other split products has further functions in immunity.

1. Levashina, E. A. *et al.* *Cell* **104**, 709–718 (2001).
2. Janssen, B. J. C. *et al.* *Nature* **437**, 505–511 (2005).
3. Janssen, B. J. C., Christodoulidou, A., McCarthy, A., Lambris, J. D. & Gros, P. *Nature* **444**, 213–216 (2006).
4. Ajees, A. A. *et al.* *Nature* **444**, 221–225 (2006).
5. Wiesmann, C. *et al.* *Nature* **444**, 217–220 (2006).
6. Reid, K. B. & Porter, R. R. *Annu. Rev. Biochem.* **50**, 433–464 (1981).
7. Thiel, S. *et al.* *Nature* **386**, 506–510 (1997).
8. Farries, T. C., Lachmann, P. J. & Harrison, R. A. *Biochem. J.* **252**, 47–54 (1988).
9. Fearon, D. T. & Locksley, R. M. *Science* **272**, 50–53 (1996).
10. Pickering, M. C. *et al.* *Nature Genet.* **31**, 424–428 (2002).
11. Haines, J. L. *et al.* *Science* **308**, 419–421 (2005).

QUANTUM PHYSICS

Information on heat

Keith Schwab

There is a fundamental quantum limit to heat flow, just as there is to electric current. This limit is independent of what carries the heat, and could also have a role in an unexpected quarter: information theory.

In the past 20 years, physicists have learnt a tremendous amount about the transport of matter and energy through devices small enough for quantum effects to come into play. One surprising fact that has emerged is that the rates of transport in such devices, expressed for example by their electronic or thermal conductance, have simple quantum-mechanical limits. On page 187 of this issue, Meschke *et al.*¹ extend this principle to heat conduction by photons. Although the result will certainly have practical ramifications for the engineering of ultra-sensitive detectors, sensors and micro-electronic refrigerators, the physics behind it hints at more fundamental truths.

The experimental trail leading to this point starts in 1988, when two groups independently demonstrated^{2,3} the quantization of electron transport through a single 'ballistic' channel in which the electron's movement is impeded only negligibly by scattering. The conductance of such a channel (the inverse of its resistance) was found to vary in discrete steps of size $2e^2/h$, where e is the electron charge and h is Planck's constant. This quantum of electrical conductance has a value of around $7.8 \times 10^{-5} \Omega^{-1}$ in conventional units. In small systems, such as a single channel, that have only one way for electrons to propagate (one 'mode of conduction'), it also represents the maximum value of the conductance.

This first demonstration of quantized electronic transport was a *tour de force* of cryogenics, microelectronics and materials science. Since then, the phenomenon has been observed in thousands of situations: in larger, mesoscopic devices, in atomic point contacts and break junctions, in single molecules, nanotubes and so on. It has become an example of what physicists love, and strive for, most: a prediction of complete generality. Without specifying any of the material details, such as the structure of electronic bands or the density, one can estimate the electronic conductance of any quantum-scale device.

Experimental evidence that the transport of thermal energy is also quantized dates back to 1999. Then, Michael Roukes and I demonstrated⁴ that the movement of heat through discrete, freely suspended mechanically vibrating channels approached a previously predicted maximum rate⁵ for each independent vibrational mode — longitudinal, transverse and torsional — of the structure. This universal rate is the quantum of thermal conductance, G_Q , and comes in units of $\pi^2 k_B^2 T/3h$. Here, k_B

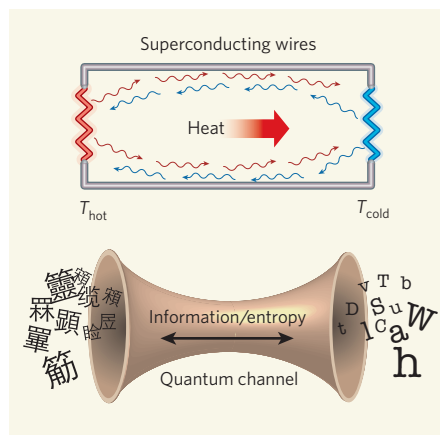


Figure 1 | Information is power. In Meschke and colleagues' experiments¹ on quantized heat transfer by photons, two resistors of different temperatures radiate thermal voltage noise. Net power and entropy flow from hot to cold at the maximum possible rate. Because entropy and information are fundamentally related⁹, this exchange of noise is analogous to classical information transfer over a quantum channel in which both parties encode information on the particles they exchange. The fact that heat transport has a universal quantum limit that is independent of the type of particle exchanged might therefore be the result of a similar principle that underlies information exchange.

is Boltzmann's constant and T is the prevailing temperature; the exact maximum rate of heat transport through a quantum device therefore increases linearly with temperature.

The mechanical vibrations that transported the heat in this case are called phonons, and are analogous to the quantized vibrations of the electromagnetic field — better known as photons. With photons, the channels of conduction are the propagating electromagnetic modes of a transmission line or of an optical waveguide such as a single-mode fibre-optic cable.

The advance made by Meschke and colleagues¹ is to demonstrate for the first time the quantized conduction of heat by photons. In their experiment, two microscopic electronic resistors exchanged heat through random thermal voltage fluctuations transmitted through two superconducting wires. With some clever use of further superconducting circuitry, the authors could switch the electrical conduction channel on and off, and thus expose the thermal connection between the two resistors brought about by the photons emitted and absorbed by them both. The authors show

that the rate of heat exchange between the two resistors in this case is given simply by G_Q .

Thus, like the quantum of electrical conductance, G_Q is very general, and independent of the nature of the material connection between two heat reservoirs. Nor does it depend on the type of particle that carries the heat: the same G_Q is the limit for the conductance of heat by electrons, phonons, photons, gravitons, you-name-it-ons. Thus, G_Q is universal in a much deeper sense^{6–8} than the quantum of electrical conductance, which depends on the quantum statistics of the particles; that is, whether they are bosons, fermions or something in between, such as the 'anyons' that crop up in certain two-dimensional systems.

A further connection that might seem somewhat surprising at first glance is that the quantum limit for heat transport is intimately related to the maximum classical information capacity of a single quantum channel (Fig. 1). This connection rests on a deep relationship between information and entropy, which was established by Claude Shannon in 1948 (ref. 9) and has been investigated by many authors over four decades¹⁰. One of the more interesting treatments⁵ made explicit the connection between maximum heating and cooling rates through both fermionic and bosonic channels on the one hand, and maximum data rates on the other. Although the maximum data rates of fermionic and bosonic channels differ, there has been speculation¹¹ that a deeper underlying principle might exist that could be used to extract a universal statement about information capacity from the truly universal nature of thermal transport in one dimension. This connection reinforces the emerging view of the basic nature of information in fundamental physics¹².

Cast in this light, the work of Meschke and colleagues¹ is of more fundamental importance than just investigating the behaviour of microscopic resistors exchanging heat. What they have demonstrated is two resistors babbling to each other using thermal voltage noise. As these resistors are near-perfect 'black-body' radiators, they emit and absorb radiation fields of maximum entropy, and so — according to Shannon's work⁹ — maximum information content. The authors have proved that this information can be carried by particles of very different natures: photons do just as well as phonons. In this spirit, I look forward to the day when a measurement of the thermal conductance through a channel of anyons in a quantum Hall fluid demonstrates the extraordinary, universal nature of G_Q . ■

Keith Schwab is in the Department of Physics, Cornell University, 618 Clark Hall, Ithaca, New York 14853, USA.
e-mail: schwab@ccmr.cornell.edu

1. Meschke, M., Guichard, W. & Pekola, J. P. *Nature* **444**, 187–190 (2006).
2. Van Wees, B. J. *et al. Phys. Rev. Lett.* **60**, 848–851 (1988).
3. Wharam, D. A. *et al. J. Phys. C* **21**, L209 (1988).
4. Schwab, K. *et al. Nature* **404**, 974–977 (2000).

5. Pendry, J. B. *J. Phys. A: Math. Gen.* **16**, 2161–2171 (1983).
6. Rego, L. G. C. & Kirczenow, G. *Phys. Rev. B* **59**, 13080 (1999).
7. Krive, I. V. & Mucciolo, E. R. *Phys. Rev. B* **60**, 1429–1432 (1999).
8. Blencowe, M. P. & Vitelli, V. *Phys. Rev. A* **62**, 052104 (2000).

9. Shannon, C. E. *Bell Syst. Tech. J.* **27**, 379–423 (1948).
10. Caves, C. M. & Drummond, P. D. *Rev. Mod. Phys.* **66**, 481–537 (1994).
11. Blencowe, M. P. *Phys. Rep.* **395**, 159–222 (2004).
12. Brassard, G. *Nature Phys.* **1**, 2–4 (2005).

EARTH SCIENCE

Isotopic hide and seek

Francis Albarède

Isotopes formed by the decay of radioactive nuclei provide evidence of how Earth was shaped in its infancy. But some decay products seem to be hidden — a finding that will revitalize a debate about Earth's interior.

It is thought that Earth formed, along with the rest of the Solar System, from the gravitational collapse of a huge cloud of gas. But what were the processes that turned the resulting ball of rubble into a planet endowed with a metallic liquid core, a thick viscous mantle and a thin continental crust? Some clues may be found by measuring the abundance of neodymium isotopes in volcanic rocks, as reported by Boyet and Carlson in *Earth and Planetary Science Letters*¹. Their results illuminate the early dynamics of Earth's interior and provide fresh insight into the structure and flow of the modern mantle.

Many short-lived radioactive elements were created in the events that accompanied the formation of the Solar System. These elements — or rather, the isotopically distinctive products of their radioactive decay — provide clues about the processes involved in planetary accretion and development. In the final stages of Earth's formation, differentiation processes distributed its constituent minerals and metals between the core, mantle and crust. This resulted in the separation of radioactive nuclides from the products of their decay, so that isotope compositions of the daughter elements in modern-day minerals attest to the separation of vapour, melts and solid phases in the newborn planet.

For example, samarium (Sm) decays to produce neodymium (Nd); both of these elements are found in mantle silicates. Variations of neodymium-isotope abundances in terrestrial rocks are useful for dating melt segregation in the mantle. Most earth scientists are familiar with the decay of ¹⁴⁷Sm into ¹⁴³Nd, a process with a half-life of 106 billion years. This decay is used as a chronometer and as a tracer of material exchange between mantle and crust. But with such a long half-life it cannot provide a detailed picture of the dawn of geological times. In contrast, the much shorter half-life of ¹⁴⁶Sm decay to ¹⁴²Nd (103 million years) is comparable to the timescale of early planetary processes and so is well suited to investigating the dynamics of newborn Earth.

So far, the ¹⁴⁶Sm–¹⁴²Nd chronometer has found its most valuable application in the study

of magma oceans — envelopes of liquid silicate that hug newborn planets. The abundances of ¹⁴²Nd in 3.8-billion-year-old rocks from Isua in Greenland are higher than those in chondrite meteorites, which are made from the same raw materials as Earth and the other rocky planets. Samarium and neodymium are 'refractory' elements, which survived the harsh conditions found in the newborn Solar System. Their relative proportions in Earth must therefore be the same as for chondrites. The 'excess' of ¹⁴²Nd, compared with chondrites, in the Isua basaltic rocks is evidence of a very early melting event (the formation of magma oceans) that concentrated ¹⁴²Nd in certain regions of the newly created mantle^{2,3}.

The modern terrestrial mantle and continental crust also display excesses of ¹⁴²Nd compared with chondrites. This geochemical feature can only be explained if the mantle is the solid residue of an ancient slurry. To balance the books, the excess of ¹⁴²Nd in these regions must be paired with a deficit in another complementary material that is derived from the liquid part of the slurry. But this ¹⁴²Nd-deficient material is conspicuously missing from the geological record. Boyet and Carlson have previously argued⁴ that this component probably took the form of a primordial crust, which sank down to the core–mantle boundary very early in Earth's history. The case for a missing reservoir, distinct from the familiar mantle, has also been made based on the abundances of the heavier neodymium isotope ¹⁴³Nd in Archaean rocks⁵ (which are more than 2.7 billion years old) and of hafnium in oceanic basalts⁶. But the ¹⁴²Nd anomalies raise an additional useful point — that the absent material was segregated in the lower mantle during the first tens of millions of years of Earth's history.

The quest for the hidden reservoir commenced with studies on oceanic basalts derived from deep mantle material⁷. But now Boyet and Carlson¹ present high-precision ¹⁴²Nd abundance data from carbonatites and diamond-bearing kimberlites — volcanic rocks that are thought most probably to originate from the deepest parts of the mantle. They report that none of these rocks shows a deviation of ¹⁴²Nd

abundances from the modern terrestrial value. This suggests that the rocks do not originate from a ¹⁴²Nd-deficient reservoir or, at the very least, that the contribution of such deep-seated material is not detectable.

The authors¹ review different interpretations of their data. They first examine the possibility that Earth's composition may not be chondritic; this could be true if atomic nuclei weren't created in a uniform distribution throughout the nebula from which the Solar System formed. A recent paper⁸ suggests that at least one process of nucleosynthesis in supernovae could lead to uneven isotopic abundances of some elements. But the isotopic distributions of nearly all the elements are the same in other planets, which makes it very difficult to justify non-chondritic abundances of samarium and neodymium for Earth.

So, if the neodymium isotope compositions of magmas reflect those of their source mantle, the conclusion from Boyet and Carlson's data is inescapable: some lower-mantle material with a ¹⁴²Nd-deficit exists, untapped by deep magmas and separate from the convective flow field of the upper mantle. This ¹⁴²Nd-deficient material may be locked up in the so-called D'' layer of the lower mantle, which sits on top of the core–mantle boundary. Or perhaps it resides in a deep reservoir resembling the abyssal layer that is proposed to exist at a depth of about 1,600 kilometres⁹, although this layer has so far evaded detection¹⁰. Of course, it could simply be that the kimberlites and carbonatites analysed by Boyet and Carlson were contaminated on their way to Earth's surface, masking the modest ¹⁴²Nd deficit inherited from their source region⁷.

Several models for the structure and dynamics of the mantle have been proposed over the years, including the theory that two separate mantle layers exist, each with its own convection patterns. This theory seemed inconsistent with emerging seismic evidence and went out of favour. But thanks to Boyet and Carlson's studies¹, layered mantle convection may now be knocking at the back door, as this could explain why deep reservoirs have become segregated from upper mantle regions. The authors have revitalized this fundamental debate — and the arguments look set to continue for years to come. ■

Francis Albarède is at Ecole Normale Supérieure de Lyon, 69007 Lyon, France.
e-mail: albarede@ens-lyon.fr

1. Boyet, M. & Carlson, R. W. *Earth Planet. Sci. Lett.* **250**, 254–268 (2006).
2. Boyet, M. *et al.* *Earth Planet. Sci. Lett.* **214**, 427–442 (2003).
3. Caro, C. *et al.* *Nature* **423**, 428–432 (2003).
4. Boyet, M. & Carlson, R. W. *Science* **309**, 576–581 (2005).
5. Chase, C. G. & Patchett, P. J. *Earth Planet. Sci. Lett.* **91**, 66–72 (1988).
6. Blichert-Toft, J. & Albarède, F. *Earth Planet. Sci. Lett.* **148**, 243–258 (1997).
7. Boyet, M., Garcia, M. O., Pick, R. & Albarède, F. *Geophys. Res. Lett.* **32**, L04306 (2005).
8. Andreasen, R. & Sharma, M. *Science* doi:10.1126/science.1131708 (2006).
9. Kellogg, L. H., Hager, B. H. & Van der Hilst, R. D. *Science* **283**, 1881–1884 (1999).
10. Vidale, J. E. *et al.* *Geophys. Res. Lett.* **28**, 859–862 (2001).

BRIEF COMMUNICATIONS

MECHANOLUMINESCENCE

Light from sonication of crystal slurries

Light flashes sparked by acoustic shock waves far outshine those created by manual crushing.

Mechanoluminescence, also known as triboluminescence or fractoluminescence, is light emission induced as a result of mechanical action on a solid^{1–3} — for example, Francis Bacon noted as long ago as 1605 that lumps of sugar emitted light when scraped⁴. Here we elicit mechanoluminescence by a new means, acoustic cavitation, and find intense luminescence and emission lines that are not generated by other mechanisms such as grinding, cleaving, rubbing, scratching, biting or thermal shock.

Acoustic cavitation is the formation, growth and implosive collapse of bubbles in a liquid irradiated with ultrasound⁵. As these bubbles collapse, extreme conditions of thousands of degrees⁶ and hundreds of atmospheres⁷ are generated within the bubble. Intense shock waves are launched outwards from the imploded bubble at velocities exceeding the speed of sound⁸. These shock waves accelerate micrometre-sized particles to high velocities and can create interparticle collisions at roughly half the speed of sound in the liquid (hundreds of metres per second)^{9,10}. Acoustic cavitation may also facilitate interparticle collisions during bubble implosion, or by rapid expansion of a bubble (at tens of metres per second) with particles attached to its surface¹¹.

About 36% of all inorganic and 19% of all organic compounds exhibit mechanoluminescence¹², but there is much we do not know about the phenomenon. When a piezoelectric crystal (such as sucrose or resorcinol) is stressed, polarization of electrical charge is induced. At fracture, the two faces of the crack will have opposite charges; as the faces separate, the local electric field created is sufficient to cause dielectric breakdown of the solid and the intervening gases^{1–3}, producing mechanoluminescence.

On fracture of sucrose in air or under nitrogen gas (N_2), the emission spectra $N_2(C^3\Pi_u - B^3\Pi_g)$ and $N_2^+(B^2\Sigma_u^+ - X^2\Sigma_g^+)$ are seen. When sucrose powder is sonicated in N_2 -sparged dodecane, the emission is the same but its intensity is much greater (Fig. 1a). The mechanoluminescence resulting from sonication is an order of magnitude greater than that from crushing sucrose in dodecane. When sucrose is crushed by hand with a glass rod, the rod moves at tenths of metres per second; however, collisions occur at hundreds of metres per second when a shock wave is launched as a result of acoustic cavitation^{9,10}.

We induced mechanoluminescence from resorcinol under various conditions (see Fig. 1b, which includes the solid-state photo-

luminescence spectrum for comparison). When resorcinol is crushed with a glass rod in air, weak $N_2(C^3\Pi_u - B^3\Pi_g)$ and $N_2^+(B^2\Sigma_u^+ - X^2\Sigma_g^+)$ lines are seen, as well as crystal luminescence; however, when resorcinol is crushed in hexadecane under helium (He) or argon (Ar), only crystal luminescence is observed. In contrast, when a slurry of resorcinol in He- or Ar-sparged hexadecane is sonicated, intense discharge lines of He and Ar gas are created, together with a weaker crystal luminescence. When hexadecane itself, or a hexadecane slurry with hydroquinone (a non-piezoelectric, non-mechanoluminescent

analogue of resorcinol), is sonicated under these conditions as controls, there is no luminescence. Other mechanoluminescent crystals that are sonicated as slurries (for example, *m*-aminophenol; see supplementary information) give an intense gas discharge and crystal luminescence, just like resorcinol.

The intensity of mechanoluminescence is directly proportional to the area of the new surfaces created during fracture¹³. The greater intensity in mechanoluminescence that we observe as a result of cavitation could be due to an increase in the frequency of mechanical action (20 kHz), which increases the rate of fracture, or to a greater intensity of collision; only the latter, however, can explain the very large increase in He or Ar line emission.

We suggest that the mechanoluminescence we observe derives from the following chain of events: acoustic cavitation induces high-velocity collisions between suspended solid particles (primarily resulting from shock waves in the liquid); this causes fracture of the crystallites and induces exceptionally high electric fields due to charge separation at the rapidly separating piezoelectric fracture faces; finally, dielectric breakdown produces an electric discharge, particularly from dissolved gases (for example, He, Ar, N_2) and at the crystal surfaces.

Nathan C. Eddingsaas, Kenneth S. Suslick

Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA
e-mail: ksuslick@uiuc.edu

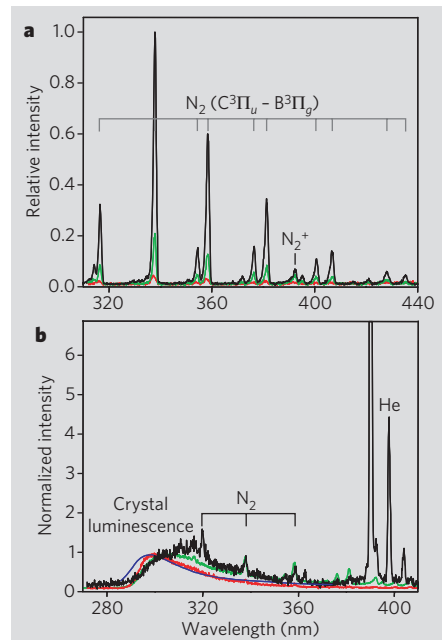


Figure 1 | Mechanoluminescent spectra from sucrose and resorcinol. **a**, Mechanoluminescence of sucrose sonicated in dodecane sparged with N_2 (black trace), crushed in air (green trace), and crushed in N_2 -sparged dodecane (red trace), all at 300 K. The emission spectra consist of the $N_2(C^3\Pi_u - B^3\Pi_g)$ and $N_2^+(B^2\Sigma_u^+ - X^2\Sigma_g^+)$ progressions. The intensity from sonication is an order of magnitude greater than from crushing by hand. **b**, Mechanoluminescence of resorcinol sonicated in hexadecane (black trace), crushed in air (green trace), crushed in He-sparged hexadecane (red trace) and its photoluminescence (blue trace) at 300 K. Strong helium (He) gas discharge is evident, together with crystal luminescence when resorcinol is sonicated in He-sparged hexadecane; N_2 is emitted because of trace N_2 contamination. No He discharge is detected from grinding in the presence of He gas.

- Chandra, B. P. in *Luminescence of Solids* (ed. Vij, D. R.) 361–389 (Plenum, New York, 1998).
- Walton, A. J. *Adv. Phys.* **26**, 887–948 (1977).
- Zink, J. I. *Acc. Chem. Res.* **11**, 289–295 (1978).
- Bacon, F. *Of the Advancement of Learning* (ed. Kitchin, G. W.) (Dent, London, 1915).
- Suslick, K. S. & Crum, L. A. in *Encyclopedia of Acoustics* (ed. Crocker, M. J.) 271–282 (Wiley, New York, 1997).
- McNamara, W. B. III et al. *Nature* **401**, 772–775 (1999).
- McNamara, W. B. III, Didenko, Y. & Suslick, K. S. *J. Phys. Chem. B* **107**, 7303–7306 (2003).
- Walton, A. J. & Reynolds, G. T. *Adv. Phys.* **33**, 595–660 (1984).
- Prozorov, T., Prozorov, R. & Suslick, K. S. *J. Am. Chem. Soc.* **126**, 13890–13891 (2004).
- Doktycz, S. J. & Suslick, K. S. *Science* **247**, 1067–1069 (1990).
- Arora, M., Ohl, C.-D. & Morch, K. A. *Phys. Rev. Lett.* **92**, 174501/1–174501/4 (2004).
- Sweeting, L. M. *Chem. Mater.* **13**, 854–870 (2001).
- Chandra, B. P. et al. *Cryst. Res. Technol.* **33**, 291–302 (1998).

Supplementary information accompanies this communication on Nature's website.

Received 20 June; accepted 12 October 2006.

Competing financial interests: declared none.
doi:10.1038/444163a

HUMAN IMMUNODEFICIENCY VIRUSES

SIV infection in wild gorillas

Chimpanzees (*Pan troglodytes troglodytes*) from west central Africa are recognized as the reservoir of simian immunodeficiency viruses (SIVcpzPtt) that have crossed at least twice to humans: this resulted in the AIDS pandemic (from human immunodeficiency virus HIV-1 group M) in one instance and infection of just a few individuals in Cameroon (by HIV-1 group N) in another¹. A third HIV-1 lineage (group O) from west central Africa also falls within the SIVcpzPtt radiation², but the primate reservoir of this virus has not been identified¹. Here we report the discovery of HIV-1 group O-like viruses in wild gorillas.

More than 30 African primate species are known to carry simian immunodeficiency viruses³, but chimpanzees are the only apes known to harbour viruses closely related to HIV-1 (refs 1, 2). We tested 378 chimpanzee and 213 gorilla fecal samples from remote forest regions in Cameroon for HIV-1 cross-reactive antibodies (for map, see supplementary information). Consistent with previous findings¹, 40 of 323 specimens from *P. t. troglodytes*, but none of 55 samples from *P. t. vellerosus*, contained antibodies reactive against HIV-1 (for details, see supplementary information).

Surprisingly, 6 of 213 fecal samples from wild-living gorillas also gave a positive HIV-1 signal. All six contained antibodies that bound the HIV-1 envelope glycoprotein (gp41) in a recombinant-based immunoassay (see supplementary information), as well as the HIV-1 Env

(gp160/gp120, gp41) and Pol (p66, p51) proteins in a confirmatory western blot (Fig. 1a).

Fecal RNA was extracted from all antibody-positive gorilla specimens and amplified by reverse transcription with polymerase chain reaction. Subgenomic *pol* (about 890 base pairs), *env* (about 390 base pairs) and/or *env/nef* (about 740 base pairs) sequences were amplified from five samples derived from three gorillas (as determined by mitochondrial and microsatellite analyses of fecal DNA; see supplementary information). Phylogenetic analysis of these sequences revealed that each of the three gorillas was infected by a distinct strain of SIV. The gorilla viruses clustered together, forming a monophyletic lineage within the SIVcpzPtt radiation that was much more closely related to HIV-1 group O than was any other known SIV (Fig. 1b, and see supplementary information). We designate this new viral lineage 'SIVgor'.

The finding of distinct but related SIVgor strains in gorillas living nearly 400 km apart suggests that, as in chimpanzees, SIV infection is endemic in gorillas. An alternative explanation could be that gorillas acquire SIV sporadically from chimpanzees, but this seems unlikely as no chimpanzee community surveyed so far, including several from habitats that overlap with those of the SIVgor-positive gorillas, harbour group O-like viruses (see supplementary information). The phylogenetic relationships shown in Fig. 1b argue that chimpanzees were the original reservoir of SIVs now found in

chimpanzees, gorillas and humans; that distinct chimpanzee communities in southern Cameroon transmitted divergent SIVcpz to humans, giving rise to HIV-1 groups M and N¹; and that chimpanzees transmitted HIV-1 group O-like viruses either to gorillas and humans independently, or to gorillas that then transmitted the virus to humans.

Gorillas are classified into two species, the western *Gorilla gorilla* and eastern *G. beringei*, which in turn are subdivided into four subspecies (*G. g. gorilla*, *G. g. diehli*, *G. b. beringei* and *G. b. graueri*)⁴. We screened the two western subspecies and found SIVgor in only *G. g. gorilla*. Additional field studies are needed to determine the overall prevalence, subspecies association, genetic diversity and natural history of SIVgor infection in wild gorillas. It will also be important to determine by what route gorillas acquired SIVgor, given that these apes are herbivores and physical encounters between gorillas and chimpanzees are believed to be rare⁵. Gorillas are hunted for food^{6,7} and medicinal use⁸, and it is possible that these practices may have been responsible for the HIV-1 group O zoonosis and could pose an ongoing risk to humans.

Fran Van Heuverswyn*, **Yingying Li†**, **Cecile Neel*‡**, **Elizabeth Bailes§**, **Brandon F. Keele†**, **Weimin Liu†**, **Severin Loul‡**, **Christelle Butel***, **Florian Liegeois***, **Yanga Bienvenue‡**, **Eitel Mpoudi Ngolle‡**, **Paul M. Sharp§**, **George M. Shaw†||**, **Eric Delaporte*#**, **Beatrice H. Hahn†**, **Martine Peeters***

*UMR145, Institut de Recherche pour le Développement and Department of International Health, University of Montpellier 1, 34394 Montpellier Cedex 5, France

e-mail: martine.peeters@mpl.ird.fr

†Departments of Medicine and Microbiology, University of Alabama at Birmingham, Alabama 35294, USA

‡Projet Prévention du Sida au Cameroun (PRESICA), BP 1874 Yaoundé, Cameroon

§Institute of Genetics, University of Nottingham, Nottingham NG7 2UH, UK

||Howard Hughes Medical Institute, Birmingham, Alabama 35294, USA

#Department of Tropical and Infectious Diseases, University Hospital of Montpellier, 34295 Montpellier, France

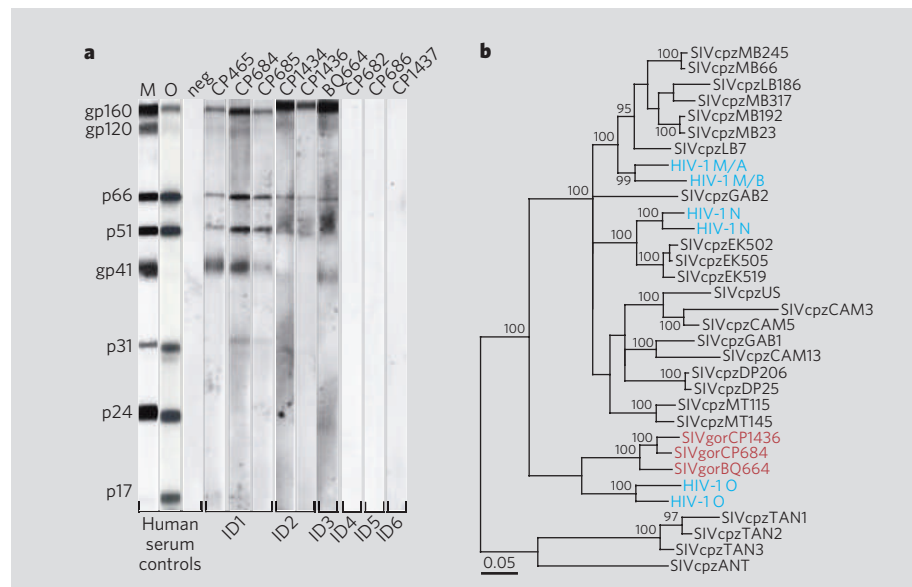


Figure 1 | Detection of SIVgor in wild gorillas. **a**, Detection of HIV-1 cross-reactive antibodies in gorilla fecal samples by western-blot analysis. Samples are numbered, with their collection site indicated (for details, see supplementary information); samples from the same individual (ID) are grouped. The relative molecular masses of HIV-1 glycoproteins (gp) and proteins (p) are indicated as a suffix (in thousands). Sera from infected (M and O) and uninfected (neg) humans are shown as controls. **b**, Evolutionary relationships of SIVgor (red), the three groups of HIV-1 (blue), and SIVcpz. Partial Pol sequences were analysed by the bayesian method; posterior probability values above 95% are shown.

1. Keele, B. F. *et al.* *Science* **313**, 523–526 (2006).
2. Sharp, P. M., Shaw, G. M. & Hahn, B. H. *J. Virol.* **79**, 3891–3902 (2005).
3. Bibollet-Ruche, F. *et al.* *J. Virol.* **78**, 7748–7762 (2004).
4. Butynski, T. M. in *Great Apes and Humans: The Ethics of Coexistence* (eds Beck, B. B. *et al.*) 3–56 (Smithsonian Institution, Washington DC and London, 2001).
5. Stanford, C. B. & Nkurunungi, J. B. *Int. J. Primatol.* **24**, 901–918 (2003).
6. Usongo, L. & Ngngue, R. *Gorilla J.* **22**, 29–31 (2001).
7. Matthews, A. & Matthews, A. *Primates* **45**, 15–24 (2004).
8. Meder, A. *Gorilla J.* **18**, 11–15 (1999).

Supplementary information accompanies this communication on Nature's website.
Received 31 August; accepted 24 October 2006.
Competing financial interests: declared none.

doi:10.1038/444164a

Great Himalayan earthquakes and the Tibetan plateau

Nicole Feld¹ & Roger Bilham¹

It has been assumed that Himalayan earthquakes are driven by the release of compressional strain accumulating close to the Greater Himalaya. However, elastic models of the Indo-Asian collision using recently imaged subsurface interface geometries suggest that a substantial fraction of the southernmost 500 kilometres of the Tibetan plateau participates in driving great ruptures. We show here that this Tibetan reservoir of elastic strain energy is drained in proportion to Himalayan rupture length, and that the consequent growth of slip and magnitude with rupture area, when compared to data from recent earthquakes, can be used to infer a ~ 500 -year renewal time for these events. The elastic models also illuminate two puzzling features of plate boundary seismicity: how great earthquakes can re-rupture regions that have already ruptured in recent smaller earthquakes and how mega-earthquakes with greater than 20 metres slip may occur at millennia-long intervals, driven by residual strain following many centuries of smaller earthquakes.

Approximately one-half of India's $36\text{--}40\text{ mm yr}^{-1}$ northward motion is absorbed by convergence of the Himalaya, one-third in contraction of the Tibetan plateau, and the remainder is distributed between Tibet and Mongolia^{1,2}. An important goal in Himalayan studies in the past decade has been to refine the Himalayan convergence rate, because this is responsible for the productivity of Himalayan earthquakes. It was anticipated that the uncertainty in the initial rate³ would steadily decrease as more GPS (Global Positioning System) data became available. Although observed velocity uncertainties have decreased as expected, calculated convergence velocities range from 14 to 20 mm yr^{-1} (Table 1), with a preferred rate in central and eastern Nepal of $19 \pm 2.5\text{ mm yr}^{-1}$ (ref. 4). In Supplementary Information we present additional GPS data from 16 points between 83°E and 87°E (Fig. 1) measured between 1991 and 2004 that further constrain the Himalayan convergence velocity.

We attempted initially to emulate the observed velocity field, as in previous studies, as the product of uniform creep on a planar dislocation beneath the plateau. This yielded a slip rate of $17 \pm 1\text{ mm yr}^{-1}$ on a 6° , N12E dipping dislocation starting at 18 km depth, consistent with earlier results (Fig. 2), but we noted that our result (and its uncertainty) was influenced by the retention or rejection of GPS data 100–300 km north of the Himalayan foothills. No abrupt transition in surface velocity distinguishes the strain field above the process zone of Himalayan seismicity from that of the Tibetan plateau, whose 17×10^{-9} strain yr^{-1} north-northeast-directed contraction and east-southeast-directed extension have been interpreted as resulting from dynamic processes responsible for inelastic, permanent deformation of the plateau³. As the inclusion of more northerly GPS points biases interpretations of Himalayan convergence velocities to higher values, we adopted a different approach, where the Himalayan and Tibetan velocity fields are considered the surface manifestation of a single deformational process related to the slip of India beneath the plateau. Our study indicates that although the region of elastic strain accumulation and release is much broader than hitherto supposed, less than one-fifth of the strain currently accumulating in the southern 500 km of the plateau participates in the earthquake cycle.

A boundary element model for interseismic deformation

We assume that frictionless aseismic slip occurs between the descending Indian plate and the overlying Tibetan plateau in an elastic half-space. We simplified the Himalayan arc as a straight line, but used a receiver-function image of the descending Indian plate⁶ as a 'cylindrical' subsurface starting geometry at its southern edge. We emulate this subsurface geometry with 14 contiguous, parallel, freely slipping boundary elements, with 64 elements along-strike⁷ extending 500 km north of the Himalaya (Fig. 2). Element dimensions are 50 km along-strike, and 2 km wide near the Himalaya increasing to 180 km in the north. We imposed an appropriate convergence velocity on the northernmost element and an overall strain contraction rate at N10E to drive the model (see Supplementary Information). We then calculated the displacement on each segment needed to minimize stress in its vicinity^{7–9} in response to these imposed loading conditions, and compared the resulting surface velocities with the observed horizontal and vertical velocities in the Himalaya and southern Tibet.

Although we tested several different northern boundaries for the model, we ultimately selected a northernmost driving segment whose southern edge lies close to where the plateau's surface converges with the Indian plate at a velocity of 21 mm yr^{-1} . This occurs ~ 550 km north of the frontal thrusts, near the Karakorum/Jiali fault system. The Himalaya advance over India at 14 mm yr^{-1} (refs 10 and 11) to 21 mm yr^{-1} (ref. 12), and the simplest assumption is that to sustain the highest rates of advance, at least 550 km of the southern plateau must participate in driving the Himalaya's southward advance. We determined that a surface convergence velocity of 21 mm yr^{-1} 550 km north of the frontal thrusts can be obtained by imposing a background strain of $-4.8 \times 10^{-8}\text{ yr}^{-1}$, and by driving the boundary element model with uniform slip of 25 mm yr^{-1} on a wide horizontal synthetic thrust fault 49 km below the plateau surface, whose southern edge lies 594 km north of the frontal thrusts. The driven dislocation is sufficiently far north to not introduce local gradients in the region of interest. To match the surface velocity field to GPS observations southward, we found it necessary to introduce minor adjustments to the depth and dip of the descending Indian plate near the transition from aseismic to seismic slip. Starting from the 'locking line' of our planar, uniform-slip model³, the best fit was obtained by

¹CIRES and Department of Geological Sciences, University of Colorado, Boulder, Colorado 80309, USA.

Table 1 | Himalayan convergence rates

Interval	Slip rate (mm yr ⁻¹)	Locking depth (km)	Dip (°)	Location	Source
1991–1995	20 ± 2 (18 ± 2*)	20 ± 4	4 ± 4	81–88° Nepal	Ref. 3
1991–1997	20 ± 1	15	3	Eastern Nepal	Ref. 36
	21 ± 2	25	4.5	Western Nepal	
	20 ± 1 (18 ± 2*)	18	5	Combined	
1991–1997	17.4 ± 0.7	9	0	Eastern Nepal	Ref. 37
	16.3 ± 0.4	12	1	Western Nepal	
	19–20	17–21	9–10	Central Nepal	
1995–2000	19	20–21	9–10	Western Nepal	Ref. 38
	14 ± 1	15	6	Northwest India	
1995–2000	19 ± 3*	ND	ND	Ladakh	Ref. 39
1997–2000	17 ± 0.9	18.3	9.5	West Himalaya	Ref. 41
1991–2000	12.2 ± 0.4	14.3	2.5	Central Himalaya	
	17.5–19	20.3	3	Eastern Himalaya	
1992–2004	19 ± 2.5	20.4	10.3	Central Nepal	Ref. 4
1992–2004	17 ± 2	18	9	Eastern Nepal	This study (planar models) ⁴²
	19 ± 3	18	6	Central Nepal	
	17 ± 1	18	9	Combined	
	21 ± 1.5	ND	ND	Central Nepal	
Holocene	≥14 ± 4	ND	ND	Kumaon Himal	Ref. 11
Holocene	≥12 ± 3	ND	ND	NW India	
Late Cenozoic	10–15*	ND	ND	73–87° Himalaya	Ref. 10
Late Cenozoic	18 ± 7*	ND	ND	73–87° Himalaya	Ref. 16

* Observed surface convergence rate, rather than estimated slip rate on dislocation. Current convergence rates are based largely on planar elastic dislocation models that derive a subsurface slip rate to emulate the surface velocities of GPS data. ND, not determined.

extending the region of aseismic slip 7.5 km further south, by reducing its depth from 18 to 15 km, and by increasing its dip from 9° to 22° N. These geometrical adjustments are permitted by current uncertainties in the geometry of the interface⁶, and the probable presence of a structural ramp beneath the Greater Himalaya. A feature of these elastic models is that they predict a broad region where the basal slip-velocity reduces to zero at the tip of the dislocation beneath and south of the Greater Himalaya. Finite element models that have included rock rheology also report a wide region where slip reduces to zero¹³. The termination of interseismic slip 83 km north of the frontal thrusts reduces the width of Himalayan ruptures from previous estimates (90–110 km) unless additional coseismic slip

extends downdip during earthquakes into the region of interseismic slip. Downdip slip here is essential either as coseismic slip or afterslip to eliminate the accumulation of a permanent slip deficit. We later compute its numerical value and spatial distribution.

The dip required by the model beneath the Greater Himalaya is in good agreement with the average dip of moderate earthquakes occurring there^{14–16}. Whereas the fit to the horizontal GPS data and vertical levelling data¹⁷ is satisfactory, predicted vertical deformation is broader than that associated with planar, uniform slip models³. Vertical GPS data in southern Tibet are insufficiently precise at present to test the predicted width of this vertical uplift^{18,19}. In contrast to the sensitivity of the model to the geometry of the descending Indian plate at depths shallower than 20 km, tests showed that the misfit in central Tibet is largely insensitive to the depth of the interface further north.

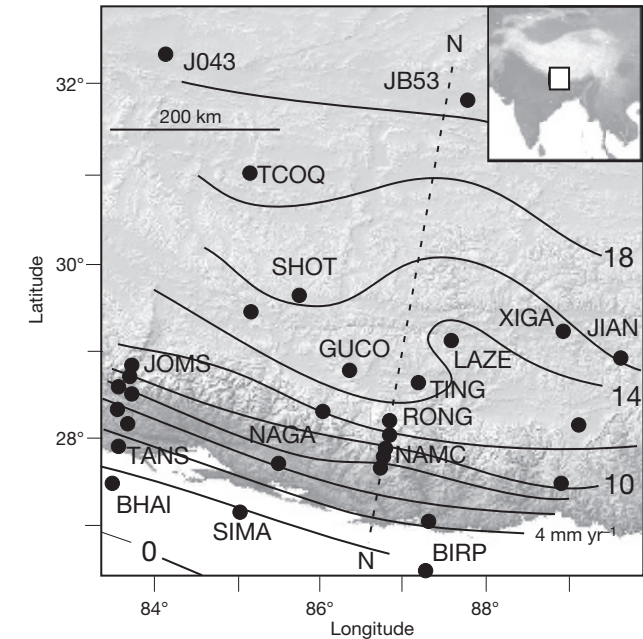


Figure 1 | The southern Tibetan plateau and Himalaya. Filled circles, GPS points; contours, the N10E GPS velocity field relative to India fixed; dashed line, the N10E velocity profile shown in Fig. 2. Inset, location of study area.

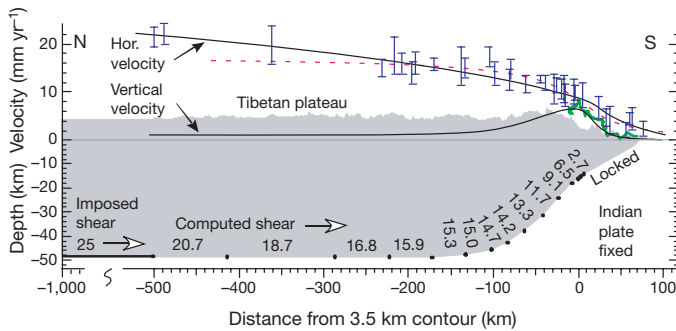


Figure 2 | Observed and synthetic present-day velocity fields for the southern Tibetan plateau. Observed fields are shown blue (horizontal GPS) and green (spirit-levelling¹⁷); synthetic fields are shown black; error bars are 1σ. Calculated interseismic segment velocities (mm yr⁻¹) and the geometry of freely slipping boundary elements are indicated at the base of plateau. The model is driven by a thrust fault in the north at 25 mm yr⁻¹ (sense of shear indicated by arrow), and a background strain contraction rate of 4.8 × 10⁻⁸ yr⁻¹ (corresponding to the mean strain rate between -500 km and the locking line). Dots indicate segment boundaries. The red dashed line shows the predicted velocity from the best-fitting planar dislocation model (see Supplementary Information). The grey shaded region represents the subsurface geometry and topography of the plateau.

Table 2 | Himalayan earthquake magnitude and slip versus rupture length

Length (km)	10	20	40	60	80	100	150	200	300	400	600	1200	2000
<i>M_w restrained</i>	6.71	6.95	7.41	7.71	7.99	8.1	8.26	8.37	8.49	8.59	8.71	8.86	9.06
<i>Max. slip (m)</i>	3.41	3.81	4.55	5.64	6.98	7.78	8.76	9.05	9.12	9.11	9.11	9.12	9.12
<i>Ave. slip (m)</i>	2.12	2.43	3.02	3.75	4.63	5.31	6.31	6.8	7.15	7.36	7.44	7.48	7.48
<i>M_w unrestrained</i>	6.72	6.96	7.42	7.77	8.05	8.18	8.36	8.5	8.66	8.77	8.9	9.13	9.28
<i>Max. slip (m)</i>	3.42	3.85	4.74	7.43	9.12	10.86	13.14	15.31	17.42	18.57	19.72	20.91	21.05
<i>Ave. slip (m)</i>	2.24	2.47	3.11	4.55	5.5	6.75	8.51	10.34	12.07	13.05	13.86	15.04	15.41
Max. 0–50 km N	0.01	0.05	0.11	1.12	1.57	2.29	3.65	5.69	8.22	9.83	11.52	13.2	13.4
Max. 50–200 km N	0	0.03	0.06	0.31	0.49	0.74	1.24	2.1	3.43	4.5	5.86	7.47	7.68

Calculations for 1,000 yr of Indo/Asian convergence at current rate. A halving in strain accumulation time (500 yr), consistent with ruptures in the past century (Fig. 4), halves the potential slip and reduces magnitudes by 0.2 M_w units. For ruptures where length $L \geq 80$ km, width w is 83 km. For ruptures where $L < 80$ km, areas are equidimensional. The first three rows in italics indicate Himalayan slip assuming that there is no coseismic slip beneath the plateau (restrained slip). The remaining five rows indicate coseismic or postseismic slip assuming that the plateau responds to the sudden southward shift in boundary conditions following Himalayan earthquakes (unrestrained slip). Unrestrained M_w magnitudes in row 4 are calculated assuming all slip is seismic. These end-member solutions form the upper and lower bounds of areas plotted in Fig. 4a and b. Maximum unrestrained trans-Himalayan coseismic displacements 0–50 km north of each rupture, and 50–200 km north of each rupture, are indicated in the last two rows.

Rupture length, earthquake magnitude and recurrence interval

We next examined the efficiency with which distributed strain stored within the southern plateau can be delivered to Himalayan earthquakes. Assuming that rupture occurs between the southernmost tip of interseismic slip beneath the plateau and the frontal thrusts, and that the slip is frictionless, the length of Himalayan ruptures determines the slip and consequently the magnitude of the earthquake. For example, if interseismic displacements were to accumulate in our model for 1,000 yr at present rates, the Himalaya would slip 21 m, were it to do so in a single massive earthquake. Assuming a rupture width of 83 km and length of 1,800 km, the moment magnitude of this improbably large earthquake would be $M_w = 9.2$ (Table 2). A 500-yr recurrence interval, however, would result in half the slip and a $M_w = 9.0$ earthquake. To calculate the slip on shorter rupture lengths, we incremented the slip velocities of our best-fitting interseismic boundary element model by 1,000 yr, and emulated an earthquake rupture by adding a freely slipping rectangular dislocation to its southernmost edge (Fig. 3). This additional rectangular region represents rupture of the Himalayan décollement, and was modelled as a 5×10 matrix of freely slipping elements, whose integrated slip and area were used to calculate M_w . In these synthetic earthquakes, we determined both the maximum and mean slip on the earthquake rupture, as well as the change of slip and strain beneath southern Tibet resulting from the occurrence of each Himalayan rupture (Table 2).

Two end-member models were considered. In the first, we calculate Himalayan slip assuming that no slip beneath the plateau accompanies rupture—that is, coseismic slip at the northern edge of the earthquake ceases abruptly at the southern edge of the region of aseismic slip. In the second, we calculate the combined change in slip on the earthquake rupture and beneath the plateau—that is, the northern edge of the earthquake rupture is unrestrained and is driven by strain relaxation in the southern plateau. This unrestrained additional slip more than doubles the maximum slip available to drive the

frontal thrusts. This down-dip slip presumably occurs as afterslip resulting in rate changes following rupture²⁰; however, the 6 June 1505 earthquake²¹ with its inferred 600-km-long rupture length was accompanied by substantial accelerations in southern Tibet, consistent with some of this down-dip slip occurring co-seismically. Our results (Table 2 and Fig. 4) demonstrate that, as expected, for earthquakes with rupture lengths shorter than the width of the Himalaya, magnitude scales with area, whereas those longer than 90 km scale with length. Similar observed, and synthetic, scaling relations have been reported for transform faults^{22–24}. Ruptures shorter than ~150 km leave strain in southern Tibet largely untapped, whereas longer ruptures drain this energy, resulting in larger slip both in the region of coseismic rupture and as increased slip down-dip from the rupture beneath the plateau (Figs 4, 5).

The renewal time for Himalayan earthquakes is unknown, because no large earthquakes have recurred repeatedly in the historical record. The slip released in recent earthquakes compared to present-day Himalayan convergence rates suggests that the renewal time is of the order of 300–1,000 yr. This estimate can be improved using the synthetic scaling law depicted in Fig. 4, by selecting a renewal time consistent with M_w versus rupture-length data from recent earthquakes. Only for the 2005 Kashmir earthquake are slip and geometry well constrained, but the convergence rate in western Kashmir is lower, and the geometry of its rupture steeper, than for earthquakes we consider here. Such data as are available are consistent with a recurrence interval of ~500 yr, and suggest that recent ruptures best fit the restrained rupture model (the lower bound of the envelopes in Fig. 4).

Whereas the maximum slip in recent earthquakes is consistent with maximum slip predicted in synthetic earthquakes with a 500-yr return time, trench excavations of surface ruptures of medieval earthquakes record slip that would require $\geq 1,000$ yr of strain accumulation^{25,26}. Minimum rupture lengths only are available for these earthquakes, but our models imply that they ruptured much longer fault planes than

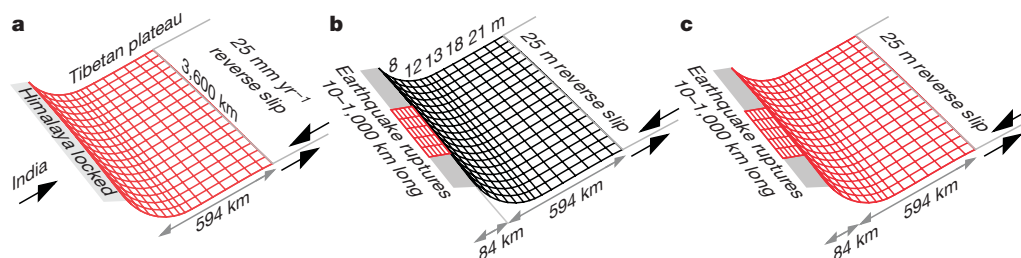


Figure 3 | Boundary element meshes used to derive synthetic slip. Red elements slip freely in response to imposed slip. **a**, Interseismic velocities calculated by driving 3,450 freely slipping elements with southerly geometry adjusted to emulate observed surface GPS data. **b**, Restrained coseismic slip. Velocities from **a** are used to derive displacements whose resulting strain field drives ruptures of different lengths in the Himalaya (Table 2). The

displacements shown correspond to a millennium of slip at current rates. **c**, Slip beneath the Himalaya changes the boundary conditions in the southern plateau, causing additional slip both on the main rupture and beneath the plateau. When fully expressed, this unrestrained slip represents maximum afterslip (Table 2 and Fig. 4).

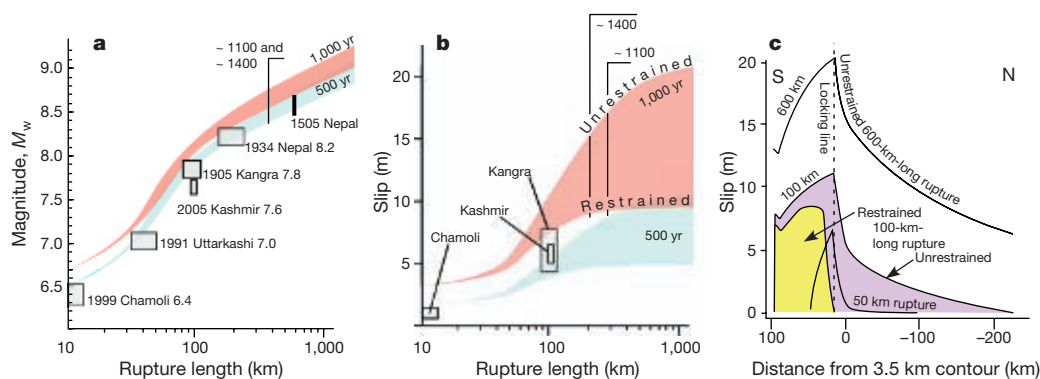


Figure 4 | Synthetic scaling laws for Himalayan earthquakes. Pink areas in **a** and **b** represent predictions for 1,000-yr intervals between earthquakes; blue areas represent predictions for 500 yr. The upper and lower bounds of these shaded areas correspond to restrained coseismic slip, and unrestrained slip, respectively. Restrained slip occurs coseismically when the interface below the plateau remains locked during the earthquake. **a**, M_w versus rupture-length curves suggest observed earthquakes recur after ~ 500 -yr intervals. **b**, Envelopes indicate minimum slip (no plateau afterslip) and maximum slip (full strain release) for Himalayan ruptures. Maximum slip in medieval earthquakes requires $>1,000$ yr strain accumulation intervals.

Open boxes are minimum estimates for historical ruptures (>300 km for about AD 1100 (ref. 26) and >200 km about AD 1400 (ref. 25)). The eastern Nepal earthquake of 6 June 1505 is believed to have ruptured 600 km along the arc²¹ and its recurrence now would release ~ 9 m of slip in a $M_w \approx 8.6$ earthquake (vertical bar in **a**). **c**, Maximum fault slip versus north/south distance as a function of rupture length. Yellow indicates restrained coseismic slip for a 100-km-long rupture (no slip north of the locking line); violet indicates complete unrestrained strain release (frictionless slip beneath the plateau). The kink in the south is caused by the surface rupture of the steeply dipping Main Boundary Fault.

recent earthquakes (Table 2). A possible explanation for these observations of large slip is that they are amplified locally by the dynamics of rupture^{23,27}, but it is more probable that long ruptures (megaquakes) at infrequent intervals are essential to occasionally reduce residual

elastic strain accumulating in the southern plateau that has been left by previous smaller earthquakes. The past 500 yr, for example, are characterized by earthquakes with rupture lengths <300 km that have inefficiently reduced cumulative elastic strain.

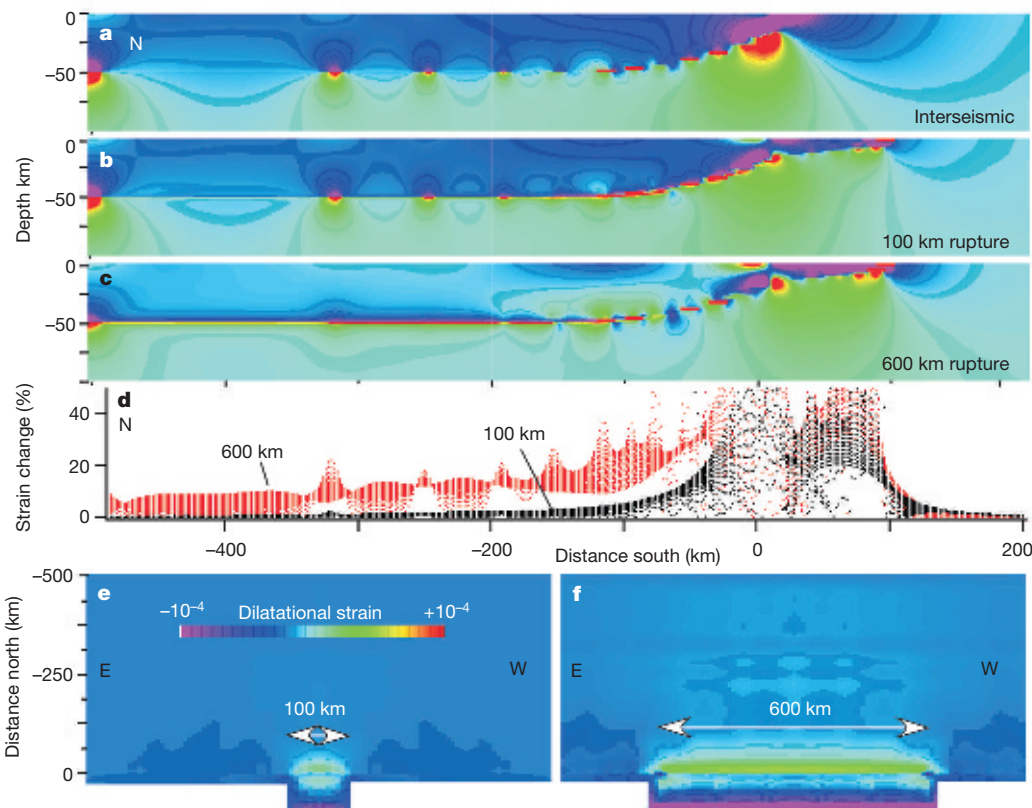


Figure 5 | Unrestrained strain changes in the southern Tibetan plateau generated by synthetic earthquakes with 100-km-long ($8 < M_w < 8.2$) and 600-km-long ($8.4 < M_w < 8.6$) ruptures. Shown are cross-sections (**a–d**) and map views (**e, f**). **a**, Pre-seismic dilatational strain associated with 1,000 yr of strain accumulation at the current rate. **b, c**, Coseismic changes caused by earthquakes with 100-km (**b**) and 600 km (**c**) rupture lengths (**e** and **f** are respective map views of surface strain changes). **d**, Calculated

percentage strain changes in the uppermost 40 km of the plateau for these two lengths of rupture. Cusps and strain concentrations in each panel are artefacts caused by junctions between boundary elements (see Fig. 2). Cross-sections **a–c** sample the centre of each rupture, but **d** samples all points shallower than 45 km throughout a ± 500 -km east–west length of the southern plateau. Colour coding for dilatational strain in **a–c** and **f** shown in **e**.

Seismic gaps versus failures in seismic gap theory

We note that our inferred renewal time applies to earthquakes of all magnitudes, which is an apparent departure from the predictions of seismic gap theory. Seismic gaps are unruptured segments of a plate boundary that slip eventually in great earthquakes. Their rupture results in an earthquake whose magnitude is characteristically proportional to the cumulative plate boundary displacement since their previous rupture. Our models suggest that once a critical strain has been attained, its release by a $M_w < 7$ or a $M_w > 9$ earthquake depends only on the length of the rupture, the growth of which in the Himalaya is presumably impeded by the presence of along-arc asperities²⁸. Insufficient earthquakes are known in the Himalaya to determine whether successive earthquakes rupture similar areas.

A second apparent departure from seismic gap theory is that our model demonstrates that a major earthquake can be followed by a great earthquake in the same location, sooner than anticipated from considerations of renewal time from plate convergence rates. Such behaviour is manifest in other convergence zones: in Chile²⁹ and in the 2004 Sumatra/Andaman earthquake³⁰. Himalayan $M_w < 8$ earthquakes drain negligible elastic energy from southern Tibet compared to those where $M_w > 8$. This can explain, for example, why the 1833 $M_w = 7.8$ Nepal earthquake was followed by the 1934 $M_w = 8.2$ Bihar/Nepal earthquake in approximately the same area after only 101 yr (ref. 31). It also implies that the 1905 Kangra earthquake region, previously considered a region relieved from an imminent damaging earthquake, could today host one of larger magnitude³².

That the elastic cycle accompanying Himalayan earthquakes must modulate the stresses that drive the long term flow behaviour of the plateau has never been questioned, but its influence in the southern-most 500 km of the plateau is unexpected. The elastic model we propose for the Tibetan plateau is a one-parameter model, driven by hypothetical reverse slip less than 500 km north of the Himalaya. This remote driving condition is not a unique requirement of the model, and similar results can be obtained using more northerly boundary conditions driven at higher rates, indicating that our conclusions are unaffected by elastic conditions in central and northern Tibet. Our traction-free lower boundary conditions at the base of the plateau are common to both dynamic³³ and kinematic models proposed for the Tibetan plateau³⁴, but unlike previous studies that incorporate realistic subsurface interface geometry, we incorporate no depth-dependent rheology, nor rate-dependent friction laws. Although the models may be refined further, the implications for earthquake scaling laws in the Himalaya are unlikely to change significantly.

The dynamic deformation of the Tibetan plateau is driven by the same forces that drive earthquakes in the Himalaya, and it is thus of some interest to estimate what fraction of present day strain accumulation in the plateau participates in the elastic earthquake cycle. Accordingly, we calculated the percentage of elastic strain released in the uppermost 40 km on a 2-km three-dimensional grid spacing throughout the southern plateau (Fig. 5d). Strain changes exceed 40% within 120 km of the Himalayan front but fall rapidly northward. For 100-km- and 600-km-long ruptures, dilatational strain changes average less than 10% and 20%, respectively, hence most of the observed interseismic strain remains to drive the dynamics and kinematics of the plateau.

The absence of surface ruptures for nineteenth- and twentieth-century central Himalayan earthquakes³⁵ suggests that their rupture lengths may have been atypically short (<200 km), and their slip correspondingly low. According to our models, the lengths of ruptures in about 1100 and 1400, which did rupture the surface, may have exceeded 400 km, with moment magnitudes exceeding $M_w = 8.6$. A ~500-yr recurrence interval for Himalayan earthquakes suggests that the western Himalaya (last mega-event in about 1400), the central Himalayan seismic gap (1505) and eastern Nepal (1100) are now sufficiently mature to sustain renewed rupture, but that the large observed slips (>20 m) of the first and last of these may require

the elapse of a further five centuries before they repeat with equal severity.

Received 26 January; accepted 25 August 2006.

- DeMets, C., Gordon, R. G., Argus, D. F. & Stein, S. Effect of recent revisions to the geomagnetic reversal time scale on estimates of current plate motions. *Geophys. Res. Lett.* **21**, 2191–2194 (1994).
- Zhang, P. et al. Continuous deformation of the Tibetan Plateau from global positioning system data. *Geology* **32**, 809–812 (2004).
- Bilham, R., Larson, K., Freymueller, J., Project Idylhim members. GPS measurements of present-day convergence across the Nepal Himalaya. *Nature* **386**, 61–64 (1997).
- Bettinelli, P. et al. Plate motion of India and interseismic strain in the Nepal Himalayan from GPS and DORIS measurements. *J. Geod.* doi:10.1007/s00190-006-0030-3 (published online 8 March, 2006).
- Houseman, G. & England, P. Crustal thickening versus lateral expulsion in the India-Asia continental collision. *J. Geophys. Res.* **98**, 12233–12249 (1993).
- Schulte-Pelkum, V. et al. Imaging the Indian Subcontinent beneath the Himalaya. *Nature* **435**, 1222–1225 (2005).
- Crouch, S. L. & Starfield, A. M. *Boundary Element Methods in Solid Mechanics* (Allen Unwin, London, 1983).
- Gomberg, J. & Ellis, M. Topography and tectonics of the central New Madrid seismic zone: Results of numerical experiments using a three-dimensional boundary-element program. *J. Geophys. Res.* **99**, 20299–20310 (1994).
- Okada, Y. Internal deformation due to shear and tensile faults in a half-space. *Bull. Seismol. Soc. Am.* **82**, 1018–1040 (1992).
- Lyon-Caen, H. & Molnar, P. Gravity anomalies, flexure of the Indian plate, and the structure, support and evolution of the Himalaya and Ganga Basin. *Tectonics* **4**, 513–538 (1985).
- Wesnousky, S. G., Kumar, S., Mohindra, R. & Thakur, V. C. Uplift and convergence along the Himalayan Frontal Thrust of India. *Tectonics* **18**, 967–976 (1999).
- Lave, J. & Avouac, J. P. Active folding of fluvial terraces across the Siwaliks Hills, Himalayas of central Nepal. *J. Geophys. Res.* **105**, 5735–5770 (2000).
- Cattin, R. & Avouac, J. P. Modelling mountain building, the seismic cycle in the Himalaya of Nepal. *J. Geophys. Res.* **105**, 13389–13407 (2000).
- Molnar, P. & Lyon-Caen, H. Fault plane solutions of earthquakes and active tectonics of the Tibetan Plateau and its margins. *Geophys. J. Int.* **99**, 123–153 (1989).
- Chen, W.-P. & Molnar, P. Seismic moments of major earthquakes and the average rate of slip in Central Asia. *J. Geophys. Res.* **82**, 2945–2969 (1977).
- Molnar, P. A review of geophysical constraints on the deep structure of the Tibetan Plateau, the Himalaya, and the Karakorum and their tectonic implications. *Phil. Trans. R. Soc. Lond. A* **326**, 33–88 (1988).
- Jackson, M. & Bilham, R. Constraints on Himalayan deformation inferred from vertical velocity fields in Nepal and Tibet. *J. Geophys. Res.* **99**, 13897–13912 (1994).
- Chen, J. Y. & Zhang, J. On the crustal movement in Quomolongma Feng and its adjacent area. *Acta Geophys. Sinica* **39**, 67–78 (1996).
- Xu, C., Jungnan Liu, W. J. & Chuang, S. GPS measurements of present day uplift in Southern Tibet. *Earth Planets Space* **52**, 735–739 (2000).
- Perfettini, H. & Avouac, J. P. Stress transfer and strain rate variations during the seismic cycle. *J. Geophys. Res.* **109**, doi:10.1029/2003JB002917 (2004).
- Ambraseys, N. & Jackson, D. A note on early earthquakes in northern India and southern Tibet. *Curr. Sci.* **84**, 570–582 (2003).
- Wells, D. L. & Coppersmith, K. L. New empirical relationships among magnitude, rupture width, rupture area, and surface displacement. *Bull. Seismol. Soc. Am.* **84**, 974–1002 (1994).
- Shaw, B. E. & Scholz, C. H. Slip-length scaling in large earthquakes: Observations and theory and implications for earthquake physics. *Geophys. Res. Lett.* **28**, 2995–2998 (2001).
- Bodin, P. & Bilham, R. 3-D geometry at transform plate boundaries: Implications for seismic rupture. *Geophys. Res. Lett.* **21**, 2523–2526 (1994).
- Kumar, S. et al. Paleoseismic evidence of great surface rupture earthquakes along the Indian Himalaya. *J. Geophys. Res.* **111**, B03304, doi:10.1029/2004JB003309 (2006).
- Lave, J. et al. Evidence for a great medieval earthquake (~1100 A.D.) in the Central Himalayas, Nepal. *Science* **307**, 1302–1305 (2005).
- Brune, J. N., Brown, S. & Johnson, P. A. Rupture mechanism and interface separation in foam rubber models of earthquakes: a possible solution to the heat flow paradox and the paradox of large overthrusts. *Tectonophysics* **218**, 59–67 (1993).
- Bollinger, L., Avouac, J. P., Cattin, R. & Pandey, M. R. Stress buildup in the Himalaya. *J. Geophys. Res.* **109**, doi:10.1029/2003JB002911 (2004).
- Cisternas, M. et al. Predecessors of the giant 1960 Chile earthquake. *Nature* **437**, 404–407 (2005).
- Bilham, R., Engdahl, E. R., Feldl, N. & Satyabala, S. P. Partial and complete rupture of the Indo-Andaman plate boundary 1847–2004. *Seismol. Res. Lett.* **76**, 299–311 (2005).
- Bilham, R. Earthquakes in India and the Himalaya: Tectonics, geodesy, and history. *Ann. Geophys.* **74**, 839–858 (2004).
- Bilham, R. & Wallace, K. Future $M_w > 8$ earthquakes in the Himalaya: Implications from the 26 Dec 2004 $M_w = 9.0$ earthquake on India's eastern plate margin. *Geol. Surv. India Spec. Publ.* **85**, 1–14 (2005).

33. England, P. & Molnar, P. Active deformation of Asia: From kinematics to dynamics. *Science* **278**, 647–650 (1997).
 34. Tapponnier, P. *et al.* Oblique stepwise rise and growth of the Tibet Plateau. *Science* **294**, 1671–1677 (2001).
 35. Bilham, R. & Ambraseys, N. Apparent Himalayan slip deficit from the summation of seismic moments for Himalayan earthquakes, 1500–2000. *Curr. Sci.* **88**, 1658–1663 (2005).
 36. Larson, K., Burgmann, R., Bilham, R. & Freymueller, J. Kinematics of the India-Eurasia collision zone from GPS measurements. *J. Geophys. Res.* **104**, 1077–1093 (1999).
 37. Burgmann, R., Larson, K. & Bilham, R. Model inversion of GPS and leveling measurements across the Himalaya: Implications for earthquake hazards and future geodetic networks. *Himal. Geol.* **20**, 59–72 (1999).
 38. Jouanne, F. *et al.* Current shortening across the Himalayas of Nepal. *Geophys. J. Int.* **157**, 1–14 (2004).
 39. Banerjee, P. & Burgmann, R. Convergence across the northwest Himalaya from GPS measurements. *Geophys. Res. Lett.* **29**, 31–34 (2002).
 40. Jade, S. *et al.* GPS measurements from the Ladakh Himalaya, India: Preliminary tests of plate-like or continuous deformation in Tibet. *Geol. Soc. Am. Bull.* **116**, 1385–1391 (2004).
 41. Chen, Q. *et al.* Spatially variable extension in southern Tibet based on GPS measurements. *J. Geophys. Res.* **109**, B09401, doi:10.1029/2002JB002350 (2004).
 42. Feldl, N. *Crustal Deformation Across the Himalaya of Central and Eastern Nepal*. MSc thesis, Univ. Colorado (2005).
- Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.
- Acknowledgements** Discussions with P. Bodin, J. Gomberg, G. King, P. Molnar and W. Szeliga, and written comments from S. Wesnousky, J.-P. Avouac and J. Freymueller, have improved the manuscript. This paper is based upon work supported by the National Science Foundation.
- Author Contributions** The 2003/4 GPS observations in Nepal were conducted by N.F. and R.B., and the planar dislocation models presented here formed the focus of N.F.'s MSc thesis at the University of Colorado.
- Author Information** Raw GPS data are available from UNAVCO (<http://www.unavco.org/>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.B. (bilham@colorado.edu).

Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*

Jean-Marc Aury^{1*}, Olivier Jaillon^{1*}, Laurent Duret², Benjamin Noel¹, Claire Jubin¹, Betina M. Porcel¹, Béatrice Ségurens¹, Vincent Daubin², Véronique Anthouard¹, Nathalie Aiach¹, Olivier Arnaiz³, Alain Billaut¹, Janine Beisson³, Isabelle Blanc³, Khaled Bouhouche⁴, Francisco Câmara⁵, Sandra Duhaucourt⁴, Roderic Guigo⁵, Delphine Gogendeau³, Michael Katinka¹, Anne-Marie Keller³, Roland Kissmehl⁶, Catherine Klotz³, France Koll³, Anne Le Mouél⁴, Gersende Lepère⁴, Sophie Malinsky⁴, Mariusz Nowacki⁴, Jacek K. Nowak⁷, Helmut Plattner⁶, Julie Poulain¹, Françoise Ruiz³, Vincent Serrano⁴, Marek Zagulski⁷, Philippe Dessen⁸, Mireille Bétermier^{3,4}, Jean Weissenbach¹, Claude Scarpelli¹, Vincent Schächter¹, Linda Sperling³, Eric Meyer⁴, Jean Cohen³ & Patrick Wincker¹

The duplication of entire genomes has long been recognized as having great potential for evolutionary novelties, but the mechanisms underlying their resolution through gene loss are poorly understood. Here we show that in the unicellular eukaryote *Paramecium tetraurelia*, a ciliate, most of the nearly 40,000 genes arose through at least three successive whole-genome duplications. Phylogenetic analysis indicates that the most recent duplication coincides with an explosion of speciation events that gave rise to the *P. aurelia* complex of 15 sibling species. We observed that gene loss occurs over a long timescale, not as an initial massive event. Genes from the same metabolic pathway or protein complex have common patterns of gene loss, and highly expressed genes are over-retained after all duplications. The conclusion of this analysis is that many genes are maintained after whole-genome duplication not because of functional innovation but because of gene dosage constraints.

Ciliates are unique among unicellular organisms in that they separate germline and somatic functions¹. Each cell harbours two kinds of nucleus, namely silent diploid micronuclei and highly polyploid macronuclei. The latter are unusual in that they contain an extensively rearranged genome streamlined for expression and divide by a non-mitotic process. Only micronuclei undergo meiosis to perpetuate genetic information; the macronuclei are lost at each sexual generation and develop anew from the micronuclear lineage.

In *Paramecium* the exact number of micronuclear chromosomes (more than 50) and the structures of their centromeres and telomeres remain unknown. During macronuclear development, these chromosomes are amplified to about 800 copies and undergo two types of DNA elimination event. Tens of thousand of short, unique copy elements (internal eliminated sequences) are removed by a precise mechanism that leads to the reconstitution of functional genes².

Transposable elements and other repeated sequences are removed by an imprecise mechanism leading either to chromosome fragmentation and *de novo* telomere addition or to variable internal deletions³. These rearrangements occur after a few rounds of endoreplication, leading to some heterogeneity in the sequences abutting the imprecisely eliminated regions³. The sizes of the resulting, acentric macronuclear chromosomes range from 50–1,000 kilo-

bases (kb) as measured by pulsed-field gel electrophoresis. Because the sexual process of autogamy results in an entirely homozygous genotype⁴, the macronuclear DNA that was sequenced was genetically homogeneous.

The *Paramecium* genome sequence

The *Paramecium* macronuclear genome sequence was established with the use of a whole-genome shotgun and assembly strategy. Paired-end sequencing of plasmid and bacterial artificial chromosome (BAC) clones provided a coverage of 13 genome equivalents (Supplementary Table S1). We assembled the sequence reads with Arachne⁵ in 1,907 contigs connected in 697 scaffolds of size greater than 2 kb, giving a total coverage of 72 Mb. Half of the assembly is contained in 64 scaffolds larger than 413 kb (the N50 size), and 96% of the assembly (69 Mb) in 188 scaffolds larger than 45 kb. A majority of the 188 largest scaffolds are complete macronuclear chromosomes, as determined by analysis of telomere repeats, and on average they contain one gap every 120 kb with a mean estimated size of 0.66 kb (Supplementary Table S2). The largest scaffold, 981 kb long, corresponds to the largest chromosome observed by pulsed-field analysis. Consistent with the estimated gap frequency and sizes, alignment of this scaffold with the finished, independently determined, reference

¹Genoscope and CNRS UMR 8030, 2 rue Gaston Crémieux CP5706, 91057 Evry, France. ²Laboratoire de Biométrie et Biologie Évolutive CNRS UMR 5558, Université Lyon 1, Villeurbanne, France. ³Centre de Génétique Moléculaire, CNRS, 91198 Gif-sur-Yvette Cedex, France. ⁴Laboratoire de Génétique Moléculaire, CNRS UMR 8541, École Normale Supérieure, 46 rue d'Ulm, 75005 Paris, France. ⁵Center for Genomic Regulation, Institut Municipal d'Investigació Mèdica, Universitat Pompeu Fabra, Pg. Marítim de la Barceloneta 37-49, Barcelona, Catalonia, Spain. ⁶Department of Biology, University of Konstanz, PO Box 5560 78457 Konstanz, Germany. ⁷Institute of Biochemistry and Biophysics, DNA Sequencing Laboratory, Polish Academy of Sciences, Pawinskiego 5a, 02-106 Warsaw, Poland. ⁸Laboratoire Génomes et Cancers, FRE 2939 CNRS, Institut Gustave Roussy, 94805 Villejuif Cedex, France.

*These authors contributed equally to this work.

sequence of that chromosome⁶ revealed only four small gaps (less than 250 bases), and less than one difference per 6 kb. If this chromosome is representative, then the present assembly contains more than 99% of the genome sequence with more than 99.95% accuracy.

The *Paramecium* macronuclear genome has a composition of 28% G + C and is devoid of highly repeated sequences (Supplementary Figure S1). The CpG dinucleotide frequency is severely depressed (0.42-fold the expected value) and this cannot be accounted for by codon usage (Supplementary Table S3). In plant and vertebrate genomes, CpG depression is thought to be a consequence of cytosine methylation, but cytosine methylation has never been observed in *Paramecium*.

The genome sequence was annotated by using a combination of *ab initio* predictions, comparative genomics and complementary DNA alignments. A specific resource of about 90,000 expressed sequence tags (ESTs) from six physiological states or developmental stages of *Paramecium* was generated for this purpose. The current protein-coding catalogue is composed of 39,642 genes (Table 1). About 55% of predicted *Paramecium* proteins have no currently referenced InterPro domain, a surprisingly high percentage that probably reflects the great evolutionary distance between *Paramecium* and other eukaryotic genomes sequenced so far. The macronuclear genome is compact (78% coding density), with extremely small intergenic regions (352 bases on average) and introns (25 bp on average). Remarkably, in contrast with other compact genomes, *Paramecium* genes contain many introns and have an intron/exon ratio closer to that of some invertebrates such as *D. melanogaster* than to that of other sequenced unicellular eukaryotes (Table 1). When looking at the ESTs mapped to each annotated gene, we found no evidence of exon skipping. Adaptation of the splicing machinery to the reduced intron sizes may be incompatible with alternative splicing.

The ciliate clade is generally considered to have evolved from an ancestor with a red algal endosymbiont that gave rise to all extant species of the chromist and alveolate groups (the 'chromalveolate' hypothesis)⁷. We failed to find any clear case of genes characteristic of the algal lineage, or of genes of chloroplastic origin (see Supplementary Information). This implies that ciliates may have lost all traces of this ancient event because they no longer need photosynthesis or other algal-associated functions. Alternatively, endosymbiosis might not have occurred at the root of the chromalveolate lineage. In this hypothesis, the plastid-derived genes observed in apicomplexan parasites may be the result of a symbiotic event that occurred after their separation from the ciliate lineage.

A surprising result of the annotation process was the prediction of 39,642 protein-coding genes, a number unlikely to be an artefact of the gene prediction procedure (see Supplementary Table S5). *P. tetraurelia* is among the most gene-rich organisms yet sequenced,

including many metazoans and plants. Using all-against-all comparisons, we observed that many of the protein-coding genes in *Paramecium* are present in closely related gene families (not shown). This redundancy may explain the high gene number.

Genome duplications

We aligned each protein against the whole proteome to identify duplicated genes (see Methods). This yields a complete picture of a whole-genome duplication (WGD) linking two by two all of the 188 large scaffolds (example in Fig. 1 and global view in the outer circle of Fig. 2).

One remarkable characteristic of this WGD is the conservation of synteny: only 8 translocations and 76 internal inversions of chromosomal segments could be deduced. We found 10 instances in which two scaffolds ending in telomeres were paired with a single, longer scaffold, indicating that the germline fragmentation regions (for example transposons) might have moved since the WGD. However, validation of this will require examination of the corresponding micronuclear regions, because DNA elimination events that lead to chromosome fragmentation can also resolve to internal deletions. A second striking feature of this recent duplication is the number of genes retained in duplicate. About 68% of the proteome is composed of two-gene families originating from this event, whereas only 32% corresponds to genes returned to a single state by the loss of one of the two paralogues. This translates into 51% of the preduplication genes still present in two copies, corresponding to 12,026 ancestral genes, a number far greater than for any previously studied WGD. The recent *Paramecium* WGD thus contrasts with the WGDs discovered in yeast, fish or higher plants in recent years, in which synteny, or duplicate gene conservation, is greatly eroded and the WGD events could not always be detected without comparison with a non-duplicated reference genome^{8–17}.

All *Paramecium* chromosomes seem to be conserved after the duplication, and the 11,451 genes lost since the WGD seem to have disappeared through local decay mechanisms acting on a single gene or small region (Supplementary Fig. S11). The fact that we could detect a wide range of decay states (Supplementary Table S8 and Supplementary Fig. S12), indicates that pseudogenization might not have occurred abruptly after the WGD event, and we cannot exclude the possibility that some of these regions are still present in the micronuclear genome but are no longer amplified during macronuclear development¹⁸. We estimate the number of recent pseudogenes present in the annotated genes to be about 1,500 (see Supplementary Information).

We inferred the ancient gene order on the basis of the actual chromosomes (see Methods), revealing that two other WGD events had occurred before the recent WGD (Fig. 2, second and third circles

Table 1 | Comparison of *P. tetraurelia* gene characteristics with selected sequenced eukaryotes

Species	Chromalveolates		Plantae		Unikonts					
	Alveolates		Red algae	Streptophytes	Amoebozoa	Opisthokonts				
	Ciliates <i>P. tetraurelia</i>	Apicomplexa <i>P. falciparum</i> *				Dictyostelids <i>D. discoideum</i> *	Ascomycetes <i>S. cerevisiae</i> §	<i>C. elegans</i>	Animals <i>D. melanogaster</i> ‡, <i>T. nigroviridis</i> ¶	<i>H. sapiens</i> *
Number of genes	39,642	5,268	5,331	25,498	12,500	5,857	21,003	14,399	23,000	22,287
Mean coding size (amino acids)	454	761	517	437	518	497	441	572	405	509
Genes with introns (%)	80	54	0.5	79	69	5.2	93	78	81	85
Mean intron size (bp)	25	179	248	170	146	207	445	1,558	568	3,365
Mean no. of introns (in spliced genes)	2.9	2.6	1.0	5.4	1.9	1.0	5.1	3.7	7.3	8.1

* Most data are from ref. 45.

† According to ref. 46.

‡ Computed with BDGP4 annotation reported at ensembl34.

§ Computed with SGD1 annotation reported at ensembl34.

|| Computed with CEL150 annotation reported at ensembl34.

¶ Estimated value from ref. 13.

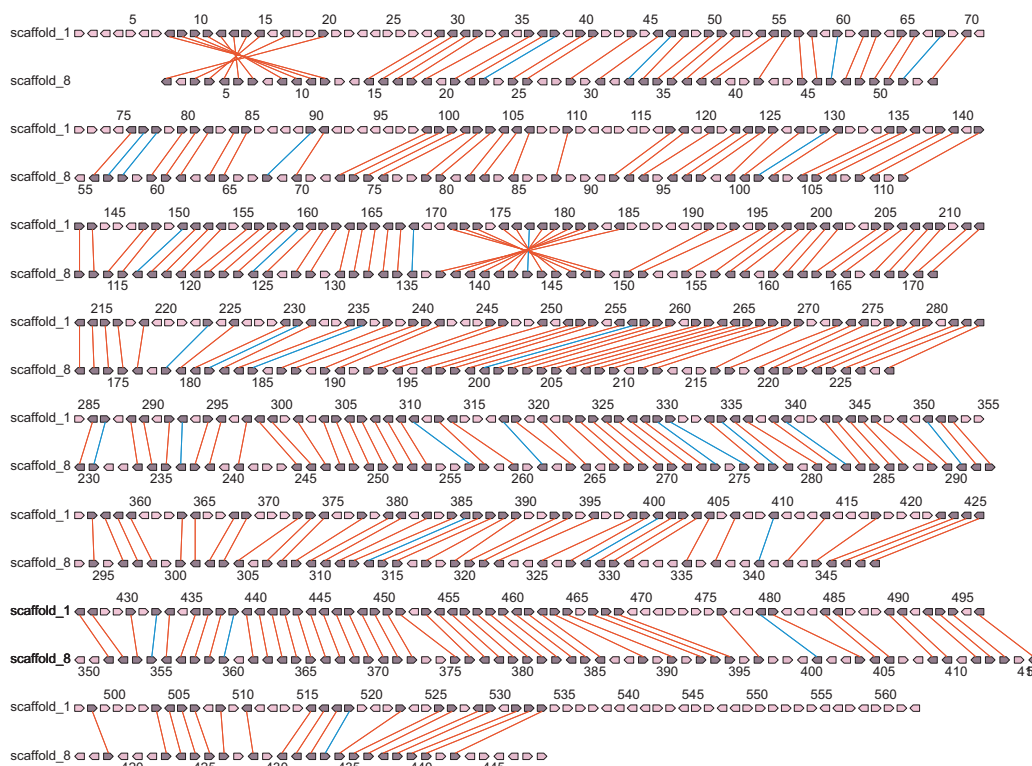


Figure 1 | Comparison of two scaffolds originating from a common ancestor at the recent WGD. Each arrow represents one gene with its direction of transcription. Brown arrows are genes with a detected paralogue, and pink arrows are genes with no paralogue. Paralogue relations

are indicated by a solid line, red for genes with a BRH match, and blue for genes with a non-BRH syntenic match. Genes at the telomeric ends have paralogues scattered between different chromosomes, probably as a result of frequent recombination.

from the outside). In comparison with the recent WGD, the number of genes conserved is smaller, as is the average similarity between paralogues (Table 2 and Fig. 3). We calculate the number of protein-coding genes in the ancestral genome before the three WGD events to be 19,552 by extrapolation of the number of genes present in the paralogons (paralogons are pairs of paralogous blocks that could be recognized as deriving from a common ancestral region). Therefore, taken together, the three WGDs have led to an approximate doubling of the gene count, explaining most of the gene number paradox in *P. tetraurelia*.

Application of the reconstruction and matching method a third time revealed a significant number of duplicated blocks with 436 duplicated genes (Fig. 2, innermost circle), and yet lower similarity between paralogues (Fig. 3a). It is therefore probable that the *Paramecium* lineage underwent a fourth, more ancient, WGD. The paralogous genes obtained in the final round of reconstruction, near detection limits, are called 'ancient duplicates', with no assumption as to their origin.

Nuclear dimorphism in *Paramecium* may have helped in tolerating WGD, because the ploidy of the expressed macronucleus is subject to physiological regulation to maintain a constant nucleocytoplasmic ratio¹⁹; it is therefore possible that a sudden duplication of the

micronuclear genome does not alter macronuclear ploidy. However, it is not clear whether this influences the recurrent appearance of WGDs, which have been found in other taxa without nuclear dimorphism^{9,10}.

Retention and evolution of duplicate genes

The nature of the duplicated genes that are retained or not after a WGD can shed light on the still debated forces shaping the genome in the wake of these drastic molecular events²⁰. Most studied WGDs are ancient and have provided clues about their long-term effects. In particular, over-retention of signalling molecules and transcription factors has been observed, linking WGD with the evolution of regulatory complexity^{21,22}. Short-term effects have not been easy to track. Very different theories on the forces that drive the retention or loss of gene duplicates include buffering for essential genes, enhancement of metabolic fluxes, protein dosage effects and rapid divergence of gene pairs^{23–28}. Because these models predict different retention outcomes according to gene classes, we analysed the predicted gene categories in *Paramecium* for their propensity for duplicate retention.

The relative expression levels of proteins involved in the same complex (or metabolic pathway; see below) must be balanced to

Table 2 | Gene content of the paralogons for each duplication

Duplication event	Genes in paralogons*	Ancestral genes† in paralogons	Ancestral genes with paralogue(s)	Ancestral genes without paralogue	Percentage of genes in the reconstruction‡
Recent WGD	35,503	35,503	24,052	11,451	90
Intermediary WGD	31,129	20,578	7,996	12,582	79
Old WGD	18,792	9,999	1,530	8,469	47
Ancient duplicates	9,735	4,830	436	4,394	25

* All genes present in a paralogon, whether a duplicated copy is retained or not.

† An ancestral gene is a gene present before a given duplication that could be represented by one or two retained paralogous copies.

‡ Percentage of genes from the annotated genome present in the ancestral blocks reconstructed from the paralogous relationships.

ensure proper stoichiometry within the complex²⁴. Thus, if duplicated genes are involved in a protein complex, the loss of one gene copy is expected to be counterselected unless it is compensated for by the upregulation of the other copy or by gene losses of other partners. To test this hypothesis we defined *Paramecium* orthologues of *Saccharomyces cerevisiae* genes involved in known complexes (Comprehensive Yeast Genome Database (CYGD) on the MIPS website (<http://mips.gsf.de/>), and ref. 29). Because we retained only clear orthologues between the two species, we expect most of the detected proteins to have ancient conserved functions and therefore a high

probability of belonging to conserved complexes. We observed a strong over-retention of genes from the recent WGD involved in known complexes (Fig. 4a). The absence of any common functional theme among the complexes analysed here makes it unlikely that there has been enrichment of specific functions. In contrast, the copy number of each protein within a complex is significantly correlated with that of the other components of the complex (*P*-value details are given in Supplementary Tables S9 and S10). We concluded that stoichiometry of the proteins constituting most complexes tends to be conserved. It has been previously shown that imbalance in the

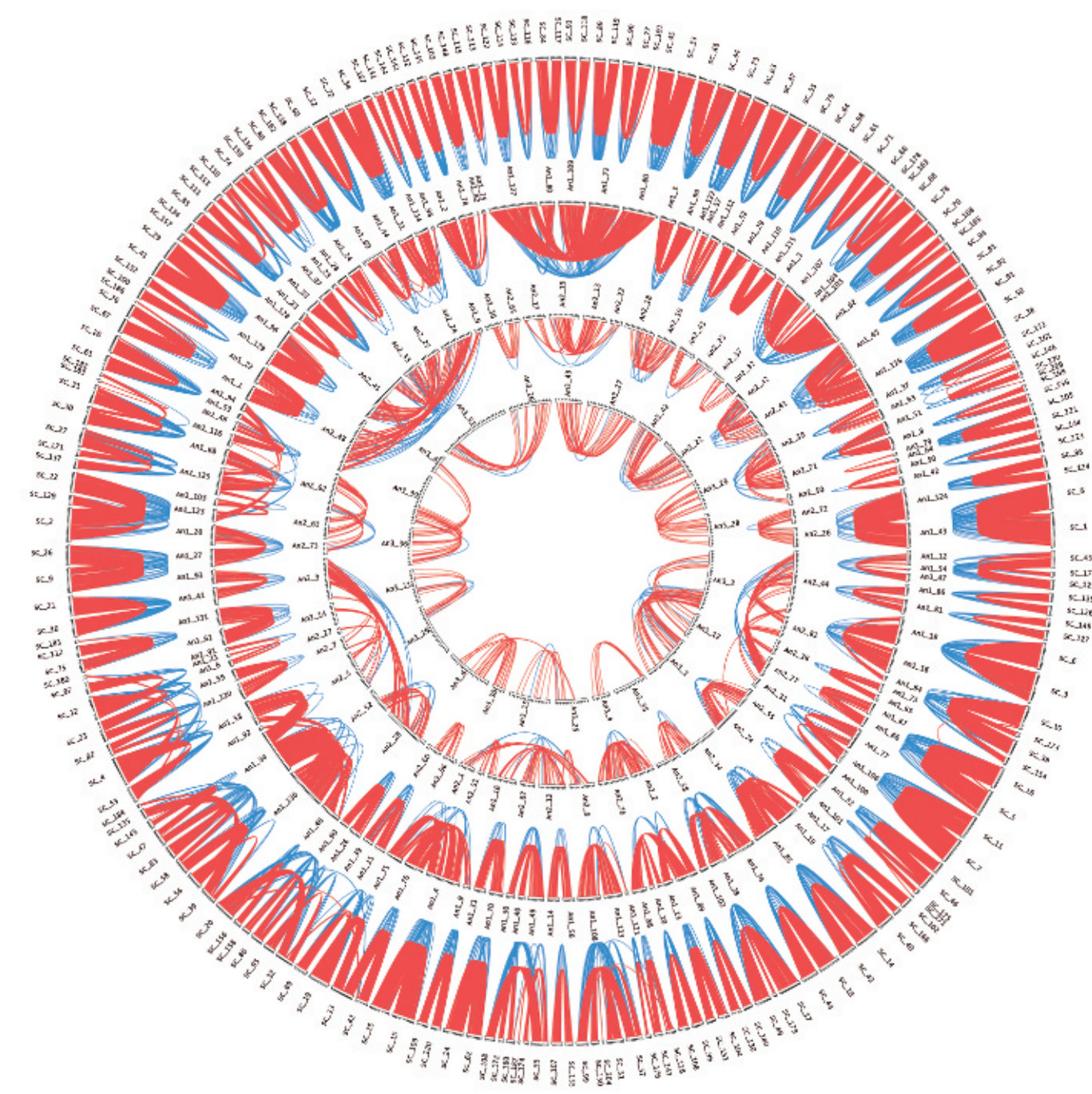


Figure 2 | Representation of the successive duplications of the *Paramecium* genome. The exterior circle displays all chromosome-sized scaffolds, and the three interior circles show the reconstructed sequences obtained by fusion of the paired sequences from each previous step. Red lines link pairs of genes with a BRH match, and blue lines link pairs of genes

with a non-BRH match that were added on the basis of syntenic position. Only the BRHs that link two genes in the same paralogon are represented. The position of an ancestral block is unrelated to the position of its constituents in the previous circle.

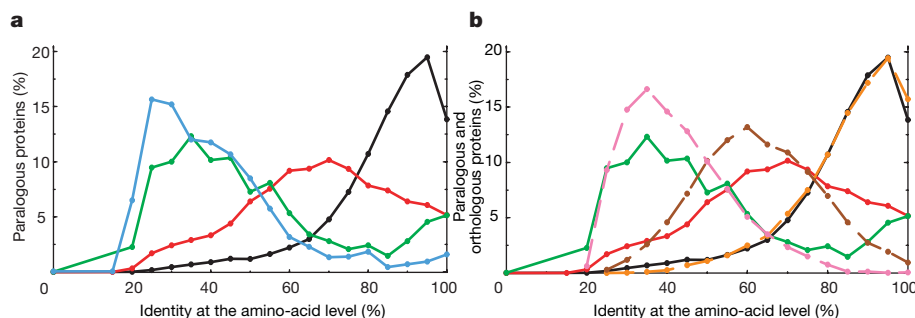


Figure 3 | Percentage identity between paralogous proteins, and comparisons with inter-species distances. **a**, Percentage identity is plotted as a function of the percentage of cases for paralogues originating from the recent (black), intermediary (red) and old (green) WGDs and from the

ancient duplicates (blue). **b**, Comparison of the three WGD plots with the distribution of orthologous matches between human and mouse (orange broken line), between human and the fish *Tetraodon* (brown broken line), and between *Paramecium* and *Tetrahymena thermophila* (pink broken line).

concentration of complex components can frequently be harmful²⁴. A key function for stoichiometry in this covariation is supported by two observations. First, the conservation of gene duplicates in a complex sometimes arises from the retention of genes originating from different duplication events, resulting in identical copy numbers for each protein (Supplementary Table S11). Second, proteins involved in many different complexes are highly retained in the recent WGD but are completely absent as retained duplicates from the old WGD (Supplementary Fig. S20). This indicates that gene loss for complex constituents might be a two-step process, in which proteins with many interactions are preferentially lost as more genes are inactivated, after a first period of preferential conservation to maintain the initial stoichiometry.

A striking result came from an analysis of central metabolism genes, which show a clear decline in retention over time. All central pathways evolve with the same trend, although at different rates (Fig. 4b). Most pathways are still significantly overamplified with respect to the recent duplication, an effect that has not been observed in WGDs analysed in other organisms^{21,30}. As with protein complexes, the genes involved in a common metabolic pathway show a tendency to covary across the duplications, behaviour that can probably be attributed to stoichiometry, because imbalance in metabolic pathways may have a similar negative effect to that in multiprotein complexes.

The loss of one paralogue would be a slow process, involving over-expression of the other copy to maintain stoichiometric constraints.

A strong correlation is observed between expression levels and retention rates of WGD paralogues across all duplications (Fig. 4c). When we consider only the proteins previously defined as associated within complexes, the level of retention is greater (Supplementary Fig. S22), showing that retention is affected both by expression levels and by association in complexes. High expression is therefore a genuine effect in duplicate retention, whether or not the protein belongs to a complex. This correlation has already been observed on a smaller scale in the yeast genome duplication³⁰, indicating that it might be a general trait of WGD.

We noticed the existence of a small group of paralogous genes that retained high expression levels, low amino-acid replacement rates and low rates of silent substitution of nucleotides. These genes also have biased codon usage (Supplementary Fig. S21), with many codons apparently optimized for high expression, because their anticodons correspond to the most abundant tRNA genes. Sequence homology allows us to propose a function for 40 of them (Supplementary Table S13). Most are involved in very basic cellular processes: ribosomal proteins, histones, cytoskeleton components, translation elongation factor 2 and succinate dehydrogenase, and many of them have conserved all eight possible copies. The resulting

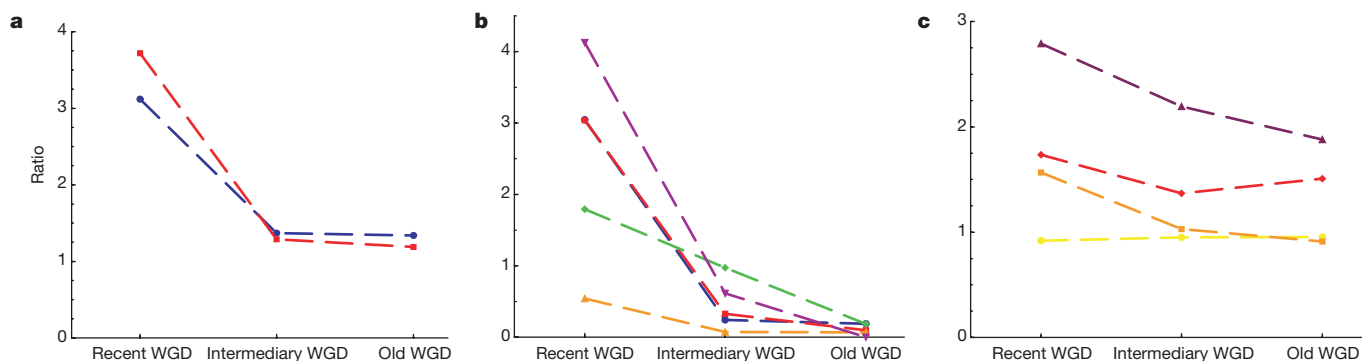


Figure 4 | Retention of duplicated genes according to biological criteria. For each curve, a value of 1 means that the retention of duplicates in one particular category is equal to the mean retention for all the genes. Values below 1 indicate under-retention and values over 1 indicate over-retention. **a**, Retention rates for proteins forming known complexes in *S. cerevisiae*. Each point represents the ratio of duplicated to non-duplicated genes for genes in complexes in *S. cerevisiae* with orthologues in *P. tetraurelia*, divided by the ratio of total duplicated genes to non-duplicated genes for each duplication. The blue curve is for orthologues of the CYGD list (<http://mips.gsf.de/>), and the red curve is for orthologues of ref. 29. **b**, Retention of gene duplicates in central metabolic pathways: blue, amino-acid

metabolism; red, carbohydrate metabolism; green, energy metabolism; orange, lipid metabolism; purple, nucleotide metabolism. Each point represents the ratio of duplicated to non-duplicated genes for genes with unambiguous EC number, divided by the ratio of total duplicated genes to non-duplicated genes for each duplication. **c**, Expression versus retention across the three WGDs. Each point represents the ratio of duplicated to non-duplicated genes for genes with a defined number of EST matches, divided by the ratio of total duplicated genes to non-duplicated genes for each duplication. Yellow, zero to two EST matches; orange, three to five EST matches; red, six to nine EST matches; purple, ten or more EST matches.

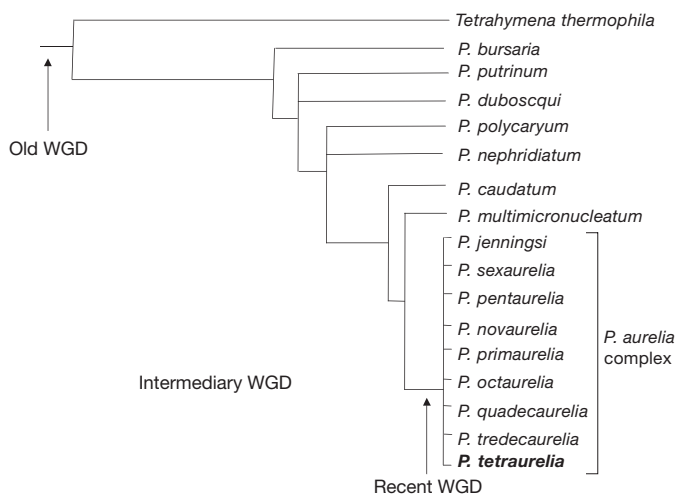


Figure 5 | Dating of genome duplication events. The phylogenetic tree of the *Paramecium* genus was adapted from ref. 47. Phylogenetic analyses indicate that the old WGD occurred before the divergence of *Paramecium* and *Tetrahymena*, and the recent WGD before the divergence of *P. tetraurelia* and *P. octaurelia*. There are currently not enough data to date the intermediary WGD.

relative enhancement of these activities should be a significant effect of WGD. Indeed, retention of duplicates for some of these genes, such as those encoding ribosome assembly proteins, is also observed in the WGDs of yeast and *Arabidopsis*^{21,22,30}.

Genome duplication and speciation

To allow dating of the intermediary and old duplications, we sampled 27 gene families for which homologues were found in *Tetrahymena thermophila* and several outgroup eukaryotes. Phylogenetic analyses indicate that the old duplication may have occurred shortly before the divergence between *Tetrahymena* and *Paramecium* lineages, whereas the intermediary duplication is specific to *Paramecium* (see Supplementary Information). The recent duplication was dated by analysing 12 gene families for which orthologues in species of the

Paramecium genus were available. Phylogenetic trees indicate, with strong bootstrap support, that the recent event occurred after the divergence of *P. tetraurelia* from *P. caudatum* and *P. multimicronucleatum* (Fig. 5). *P. tetraurelia* belongs to a complex of 15 sibling species (the *Paramecium aurelia* complex)³¹. Data from other species of the *P. aurelia* complex indicate that the recent WGD occurred before their divergence from *P. tetraurelia*, which is indicative of a possible link between genome duplication and speciation. Indeed, after a genome duplication, many gene losses occur independently in different populations. This process is therefore expected to lead rapidly to reproductive isolation by Dobzhansky–Muller incompatibility³². Interestingly, species of the *P. aurelia* complex are extremely similar both morphologically and in terms of ecological environment (and hence they were initially thought to correspond to a single species³¹). We therefore propose that the explosion of speciation events that gave rise to the *P. aurelia* complex is not the result of adaptive evolutionary events (for example the colonization of new ecological niches) but the neutral consequence of the genome duplication³².

WGDs in genome evolution

Although there is no current consensus on gene fates after a WGD, many current models consider that gene loss is rapid after the duplication, and that functional divergence (either by sub-functionalization or neo-functionalization) can explain most of the retention pattern. The *Paramecium* data point to another model, which is summarized in Fig. 6. Some genes, representing a fraction of the genome, probably quickly lose their duplicated state, because they can never be detected as duplicates in any of the three WGDs. Duplication of these genes may provide no advantage or may even be counter-selected. However, most of the gene duplicates did not go through this rapid elimination. Their maintenance may be driven at least partly by the maintenance of high expression and stoichiometric constraints. Indeed, many of the duplicates of the recent WGD are functionally redundant. Patterns of retention of more ancient WGDs indicate that most of these duplicates will be progressively lost. Yet the distribution of K_a/K_s ratios shows that almost all recent duplicates are under strong purifying selection (Supplementary Fig. S18), indicating that deleterious mutations affecting the coding sequence of one

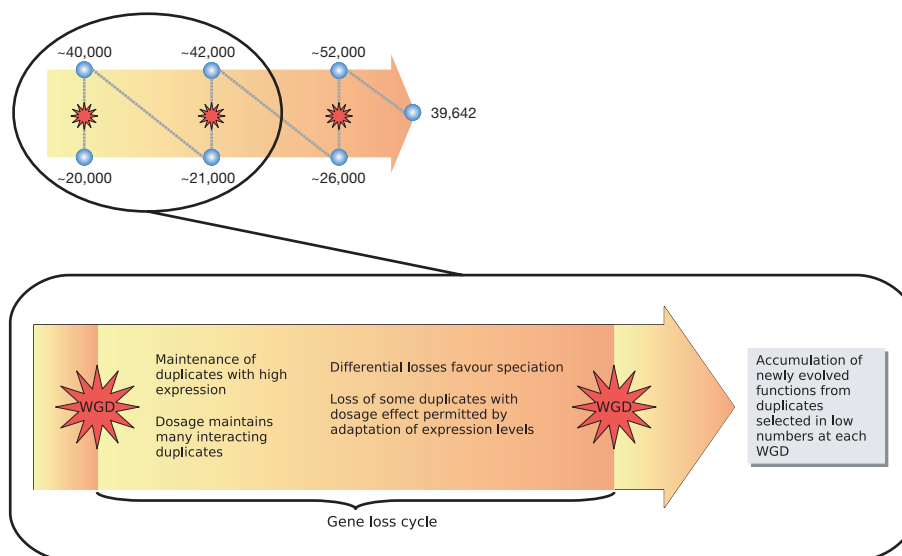


Figure 6 | A model for gradual gene loss after a WGD deduced from differential retention of genes in *Paramecium*. Top: evolution of gene number through the three WGD events. Estimates of gene numbers at the time of the duplications were made under the simplifying assumption that further gene loss affected only the new duplicates. Bottom: focus on an inter-WGD period. Just after WGD, genes subject to dosage constraints (that is, genes in complexes or metabolic pathways) and highly expressed genes are

preferentially retained as duplicates. Differential loss of paralogues in subpopulations will create barriers to reproduction, favouring speciation. Gradually, the duplicates will erode as the dosage effects are compensated for by new mutations. This massive retention of duplicates across a long evolutionary time enables selection to retain a few new genes with new derived functions.

gene are not masked by the expression of the other copy. As gene regulation evolves to permit loss without affecting dosage, most genes then gradually resolve to a single state. Another model for duplicate retention favours a buffering effect, in which genes are retained in two copies to provide robustness against mutation for important functions²³, which would be especially important for a homozygous population. It is worth noting that many plants and *Paramecium* can undergo self-fertilization and that natural populations of *P. tetraurelia* are probably homozygous most of the time. This model is consistent with high retention for basic physiological processes in *Paramecium*. However, it is difficult to consider buffering as the main effect of WGD, because it does not explain why so many genes have been retained in many more than two copies.

The appearance of new derived functions may be a secondary product of this basic model. Entire gene modules are retained initially as duplicates with redundant functions, but some of them will evolve to derived ones if there is no need other than stoichiometry for maintaining the redundancy. One example of neo-functionalization is provided by δ - and η -tubulins, paralogues of the old WGD, which have distinct functions in the assembly (δ -tubulin³³) or duplication (η -tubulin³⁴) of ciliary basal bodies. η -Tubulin is known only in ciliates, whereas δ -tubulin is conserved in eukaryotes that perpetuate the centriolar structure.

To gain insight into the acquisition of new functions through WGD, we looked for cases of marked sequence divergence between paralogues, because the neo-functionalization model predicts that the two copies will exhibit asymmetric rates of molecular evolution. After correction for multiple testing, we find that 10.9% of recent duplicates and 16.2% of intermediary duplicates show asymmetric evolutionary rates. This difference is significant ($\chi^2 = 7.55$, $P < 0.01$; see Supplementary Information) and can probably be attributed to the progressive nature of duplicate gene loss: a larger proportion of functionally redundant genes remain among recent duplicates than among the intermediary ones.

Among the 62 intermediary duplicates with asymmetric evolutionary rate, we noticed that the rapidly evolving copy tends to be less retained after the recent WGD than the slowly evolving one (26% versus 66%; $\chi^2 = 20$, $P < 10^{-4}$). This indicates that neo-functionalized copies might be more prone to pseudogenization at subsequent WGDs.

The sub-functionalization model predicts that a gene that has been preserved by sub-functionalization at a given WGD is less likely to be retained in two copies at a subsequent WGD³⁵. We observed the opposite pattern: 57% of genes already retained in the intermediary duplication are retained after the recent WGD, whereas only 47% of genes not retained after the intermediary duplication are retained after the recent WGD ($P < 10^{-4}$). Because this might simply reflect different retention rates between functional classes, only genes from the same functional class were considered. We selected families resulting from the ancient duplication for which one member has been retained in two copies after the intermediary WGD but the other has not ($n = 343$ families). In agreement with the prediction of the sub-functionalization model, the frequency of retention after the recent WGD was lower for genes retained after the intermediary WGD than for genes that had lost their duplicate (60% versus 67%; $\chi^2 = 4.58$, $P < 5\%$). Although significant, this difference is not very strong, which is consistent with the fact that sub-functionalization is an unlikely evolutionary pathway in species with large population sizes³⁵.

Conclusion

The *Paramecium* duplications have many traits that render them particularly attractive for studying the fates of duplicate genes. Both the high level of gene retention and the strong conservation of synteny helped to define an unprecedentedly large collection of paralogues, with high confidence in their exact origin. In particular, the most recent WGD proved informative on the early evolution of

paralogues. We predict a pattern of gene loss that is consistent with progressive decay, acting predominantly at the gene level. The initial maintenance of a complete double set of genes is therefore probably an essential difference between WGD and other large-scale duplications.

We can see parallel evolution for whole protein complexes, metabolic pathways, and known interacting protein groups. All these observations are consistent with a central function of dosage constraints on the evolution of gene duplicates after WGD. We propose a new model for WGD evolution, in which expression and stoichiometry are important in the retention of many duplicate genes, that can thereafter evolve in a few cases towards new derived functions. WGD is therefore fundamentally different from single-gene duplication. Indeed, the gene categories that participate in these two duplication mechanisms are significantly different²². The fact that WGDs have been postulated or demonstrated at important diverging points in eukaryotic evolution^{28,36–38} makes plausible the direct link between these amplifications and the appearance of some eukaryote-specific novelties. *Paramecium* supports different mutational systems for genetic analysis, including the use of RNA-mediated interference, that can be implemented at the genome scale. We expect that the present study will lead to further work exploring the functional fate of WGDs in detail.

METHODS

Sequencing and assembly. The whole-genome shotgun and assembly method is described in detail in Supplementary Information.

Gene annotation. Protein-coding genes were predicted by combining *ab initio* models, *Paramecium* cDNA alignments, and alignments of proteins and genomic DNA from other species. The integration of the data was realized using GAZE³⁹. Details are given in Supplementary Information.

Genome duplication. A global all-against-all comparison of every predicted protein was performed with the Smith–Waterman algorithm (e value < 0.1) and a set of 12,504 best reciprocal hits (BRHs) based on alignment scores. Each scaffold was scanned with a sliding window of BRHs with window size (w) of 10, and 60% of genes were paired (parameter p). Paralogous blocks were obtained with this windowing strategy and by merging contiguous windows associated with a common target scaffold. In the next step, 1,477 non-BRH syntenic matches found among the ten best matches of any protein with no BRH were added to the paralogue list. These paralogons were fused to create a single molecule with the assumption that every gene without a paralogue corresponds to a gene whose paralogue was lost after the duplication. This creates a succession of genes approximating the gene order of the pre-duplication region, which we called an ancestral block.

To describe the older duplications, the same approach was iterated three times with the following parameters: intermediary WGD, $w = 10$ and $p = 41\%$; old WGD and ancient duplicates, $w = 20$ and $p = 30\%$. The paralogous links for each duplication are presented in Supplementary Tables S17–S20, and the composition of the ancestral blocks is given in Supplementary Table S21. Figure 2 was constructed using Circos (<http://mkweb.bcgsc.ca/circos>).

Evolution of genes and non-coding sequences. K_1 and K_2 values were calculated on the entire set of paralogues for each WGD, using the codeml software from the PAML package⁴⁰.

Analysis of metabolic pathways. We used PRIAM⁴¹ to identify 5,617 genes containing at least one Enzyme Commission (EC) number. They were remapped on to metabolic pathways by using the KEGG database⁴².

Protein complexes. We remapped *Saccharomyces cerevisiae* complexes (ref. 29 and <http://mips.gsf.de/>) on to the *Paramecium* proteome by using orthologous links (see Supplementary Information). A protein–protein interaction was considered conserved if each protein had an orthologue in the *Paramecium* genome, and a protein complex was considered conserved if it contained at least one conserved protein–protein interaction. From the 1,602 complexes of the MIPS catalogue and the 422 complexes from ref. 29, we found respectively 443 and 61 conserved complexes in the paralogons of the old WGD.

Phylogeny of the WGDs. Phylogenetic trees were inferred from protein sequence alignments by maximum likelihood. We used Phyml⁴³ under the JTT model of protein evolution⁴⁴, with site-to-site rate variation modelled on a discrete γ distribution (four categories, shape parameter α and proportion of invariable sites estimated from the data). The robustness of the phylogenetic inference was estimated by bootstrap (500 replicates). Phylogenetic trees are

available as Supplementary Information. Data on the *Tetrahymena* proteins were from The Institute for Genomic Research (<http://www.tigr.org/tdb/e2k1/ttg/>).

Received 4 May 2006; accepted 31 August 2006.

Published online 1 November 2006.

- Jahn, C. L. & Klobutcher, L. A. Genome remodeling in ciliated protozoa. *Annu. Rev. Microbiol.* **56**, 489–520 (2002).
- Betermier, M. Large-scale genome remodelling by the developmentally programmed elimination of germ line sequences in the ciliate *Paramecium*. *Res. Microbiol.* **155**, 399–408 (2004).
- Le Mouel, A., Butler, A., Caron, F. & Meyer, E. Developmentally regulated chromosome fragmentation linked to imprecise elimination of repeated sequences in paramecia. *Eukaryot. Cell* **2**, 1076–1090 (2003).
- Sonneborn, T. M. In *Handbook of Genetics* (ed. King, R.) 469–594 (Plenum, New York, 1974).
- Jaffe, D. B. *et al.* Whole-genome sequence assembly for mammalian genomes: Arachne 2. *Genome Res.* **13**, 91–96 (2003).
- Zagulski, M. *et al.* High coding density on the largest *Paramecium tetraurelia* somatic chromosome. *Curr. Biol.* **14**, 1397–1404 (2004).
- Bhattacharya, D., Yoon, H. S. & Hackett, J. D. Photosynthetic eukaryotes unite: endosymbiosis connects the dots. *Bioessays* **26**, 50–60 (2004).
- Vision, T. J., Brown, D. G. & Tanksley, S. D. The origins of genomic duplications in *Arabidopsis*. *Science* **290**, 2114–2117 (2000).
- Blanc, G., Hokamp, K. & Wolfe, K. H. A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. *Genome Res.* **13**, 137–144 (2003).
- Bowers, J. E., Chapman, B. A., Rong, J. & Paterson, A. H. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* **422**, 433–438 (2003).
- Dietrich, F. S. *et al.* The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science* **304**, 304–307 (2004).
- Dujon, B. *et al.* Genome evolution in yeasts. *Nature* **430**, 35–44 (2004).
- Jaillon, O. *et al.* Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* **431**, 946–957 (2004).
- Kellis, M., Birren, B. W. & Lander, E. S. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**, 617–624 (2004).
- Simillion, C., Vandepoele, K., Van Montagu, M. C., Zabeau, M. & Van de Peer, Y. The hidden duplication past of *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **99**, 13627–13632 (2002).
- Wolfe, K. H. & Shields, D. C. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**, 708–713 (1997).
- Yu, J. *et al.* The genomes of *Oryza sativa*: a history of duplications. *PLoS Biol.* **3**, e38 (2005).
- Dubrana, K. & Amar, L. Programmed DNA under-amplification in *Paramecium primaurelia*. *Chromosoma* **109**, 460–466 (2000).
- Berger, J. D. & Schmidt, H. J. Regulation of macronuclear DNA content in *Paramecium tetraurelia*. *J. Cell Biol.* **76**, 116–126 (1978).
- Andalis, A. A. *et al.* Defects arising from whole-genome duplications in *Saccharomyces cerevisiae*. *Genetics* **167**, 1109–1121 (2004).
- Blanc, G. & Wolfe, K. H. Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell* **16**, 1679–1691 (2004).
- Maere, S. *et al.* Modeling gene and genome duplications in eukaryotes. *Proc. Natl Acad. Sci. USA* **102**, 5454–5459 (2005).
- Chapman, B. A., Bowers, J. E., Feltus, F. A. & Paterson, A. H. Buffering of crucial functions by paleologous duplicated genes may contribute cyclicity to angiosperm genome duplication. *Proc. Natl Acad. Sci. USA* **103**, 2730–2735 (2006).
- Papp, B., Pal, C. & Hurst, L. D. Dosage sensitivity and the evolution of gene families in yeast. *Nature* **424**, 194–197 (2003).
- Kondrashov, F. A., Rogozin, I. B., Wolf, Y. I. & Koonin, E. V. Selection in the evolution of gene duplications. *Genome Biol.* **3**, RESEARCH0008 (2002).
- Lynch, M. & Katju, V. The altered evolutionary trajectories of gene duplicates. *Trends Genet.* **20**, 544–549 (2004).
- Veitia, R. A. Paralogues in polyploids: one for all and all for one? *Plant Cell* **17**, 4–11 (2005).
- Ohno, S. *Evolution by Gene Duplication* (Allen & Unwin, London, 1970).
- Gavin, A. C. *et al.* Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631–636 (2006).
- Seoighe, C. & Wolfe, K. H. Yeast genome evolution in the post-genome era. *Curr. Opin. Microbiol.* **2**, 548–554 (1999).
- Coleman, A. W. *Paramecium aurelia* revisited. *J. Eukaryot. Microbiol.* **52**, 68–77 (2005).
- Scannell, D. R., Byrne, K. P., Gordon, J. L., Wong, S. & Wolfe, K. H. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* **440**, 341–345 (2006).
- Garreau de Loubresse, N., Ruiz, F., Beisson, J. & Klotz, C. Role of delta-tubulin and the C-tubule in assembly of *Paramecium* basal bodies. *BMC Cell Biol.* **2**, 4 (2001).
- Ruiz, F. *et al.* The *SM19* gene, required for duplication of basal bodies in *Paramecium*, encodes a novel tubulin, η -tubulin. *Curr. Biol.* **10**, 1451–1454 (2000).
- Force, A. *et al.* Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531–1545 (1999).
- Dehal, P. & Boore, J. L. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.* **3**, e314 (2005).
- Postlethwait, J., Amores, A., Cresko, W., Singer, A. & Yan, Y. L. Subfunction partitioning, the teleost radiation and the annotation of the human genome. *Trends Genet.* **20**, 481–490 (2004).
- De Bodd, S., Maere, S. & Van de Peer, Y. Genome duplication and the origin of angiosperms. *Trends Ecol. Evol.* **20**, 591–597 (2005).
- Howe, K. L., Chothia, T. & Durbin, R. GAZE: a generic framework for the integration of gene-prediction data by dynamic programming. *Genome Res.* **12**, 1418–1427 (2002).
- Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556 (1997).
- Claudel-Renard, C., Chevalet, C., Faraut, T. & Kahn, D. Enzyme-specific profiles for genome annotation: PRIAM. *Nucleic Acids Res.* **31**, 6633–6639 (2003).
- Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **32**, D277–D280 (2004).
- Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704 (2003).
- Jones, D. T., Taylor, W. R. & Thornton, J. M. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**, 275–282 (1992).
- Eichinger, L. *et al.* The genome of the social amoeba *Dictyostelium discoideum*. *Nature* **435**, 43–57 (2005).
- Matsuzaki, M. *et al.* Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* **428**, 653–657 (2004).
- Fokin, S. I. *et al.* Morphological and molecular investigations of *Paramecium schewiakoffii* sp. nov. (Ciliophora, Oligohymenophorea) and current status of distribution and taxonomy of *Paramecium* spp. *Eur. J. Protist.* **40**, 225–243 (2004).
- Keeling, P. J. *et al.* The tree of eukaryotes. *Trends Ecol. Evol.* **20**, 670–676 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Y. Bourguignon for help in statistical analyses. This work has been funded by the Consortium National de Recherche en Génomique, the CNRS, the Polish Ministry of Science and the Agence Nationale de la Recherche.

Author Information The final assembly and annotation are deposited in the EMBL/Genbank/DBJ databases under accession numbers CT867985–CT868681. An annotation browser and further information on the project are available from <http://www.genoscope.cns.fr/paramecium> and <http://paramecium.cgm.cnrs-gif.fr>. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.W. (pwinker@genoscope.cns.fr).

Crystal structure of a rhomboid family intramembrane protease

Yongcheng Wang¹, Yingjiu Zhang^{1†} & Ya Ha¹

Escherichia coli GlpG is an integral membrane protein that belongs to the widespread rhomboid protease family. Rhomboid proteases, like site-2 protease (S2P) and γ -secretase, are unique in that they cleave the transmembrane domain of other membrane proteins. Here we describe the 2.1 Å resolution crystal structure of the GlpG core domain. This structure contains six transmembrane segments. Residues previously shown to be involved in catalysis, including a Ser–His dyad, and several water molecules are found at the protein interior at a depth below the membrane surface. This putative active site is accessible by substrate through a large ‘V-shaped’ opening that faces laterally towards the lipid, but is blocked by a half-submerged loop structure. These observations indicate that, in intramembrane proteolysis, the scission of peptide bonds takes place within the hydrophobic environment of the membrane bilayer. The crystal structure also suggests a gating mechanism for GlpG that controls substrate access to its hydrophilic active site.

The concept of intramembrane proteolysis emerged from studies on sterol regulatory element-binding protein (SREBP) and amyloid precursor protein (APP)^{1,2}. The activation of SREBP requires it to be cleaved by S2P within its first transmembrane segment^{3,4}. To generate amyloid β -peptide associated with Alzheimer’s disease, the transmembrane domain of APP needs to be cleaved by γ -secretase⁵. Now we know that many proteins undergo similar scission within their transmembrane domains, and such processing is important to their biological function (for example, see refs 6, 7). In addition to S2P and γ -secretase, new members have also been added to the list of enzymes that catalyse this reaction (for example, see refs 8–11). The rhomboid family of proteases is among the recent additions^{12,13}; they are not related to S2P or γ -secretase by amino acid sequence.

Rhomboid proteases are evolutionarily widespread^{14–16}. *Drosophila* Rhomboid-1, the most characterized in the family¹⁷, regulates epidermal growth factor receptor signalling by cleaving the transmembrane domain of Spitz, the principal ligand for the receptor in flies, and promoting its release from signal-sending cells¹². Yeast rhomboid Pcp1 regulates mitochondria membrane remodelling¹⁸. The human homologue of Pcp1, PARL, regulates cytochrome *c* release during apoptosis¹⁹. Rhomboid family member AarA of *Providencia stuartii* is probably responsible for producing an extracellular quorum-sensing signal¹³. In *Toxoplasma gondii*, rhomboid protease may have a role in parasite invasion²⁰.

S2P, presenilin (the catalytic component of γ -secretase) and rhomboid are all integral membrane proteins^{5,21–23}. On the basis of sequence and mutagenesis analyses, it has been proposed that S2P, presenilin and rhomboid use hydrolytic mechanisms similar to those used by soluble metallo-, aspartyl and serine proteases, respectively^{4,12,24}. Although the location of their catalytic residues in hydrophobic regions of the protein sequence seems to suggest that their active sites are embedded within the membrane (for a review, see refs 22, 23), how these proteases catalyse peptide hydrolysis in lipid was poorly understood. Here we describe the crystal structure of a bacterial rhomboid, which is the first atomic-scale representation of any intramembrane protease, and discuss its implications for the mechanism of intramembrane proteolysis.

Overall structure

All rhomboid proteases, including *E. coli* GlpG, have a core catalytic domain of six characteristic hydrophobic segments^{12,15,21,25} (Fig. 1a; see also Supplementary Figs 1 and 2). They cleave type 1 transmembrane substrates a few residues inside of the membrane from the extracellular side^{16,21,26,27} (Fig. 1b). We have solved the structure of the GlpG core domain by X-ray crystallography at 2.1 Å resolution (Table 1; see also Supplementary Table 1 and Supplementary Figs 3 and 4).

As predicted^{12,21,25}, GlpG is composed of six transmembrane helices (S1–S6 in Fig. 1). Two features of the structure, however, were not anticipated. The first is an internal cavity. The amino terminus of the central helix S4 is about 10 Å below the membrane surface (Fig. 2b), which creates an aqueous cavity inside of the protein that opens to the extracellular side. The second unanticipated feature is a membrane-embedded loop, L1. All rhomboid protease family members have a large stretch of sequence between S1 and S2 that corresponds to this loop. The position of L1 in relation to the rest of the protein structure suggests that L1 is normally inserted in the outer leaflet of the lipid

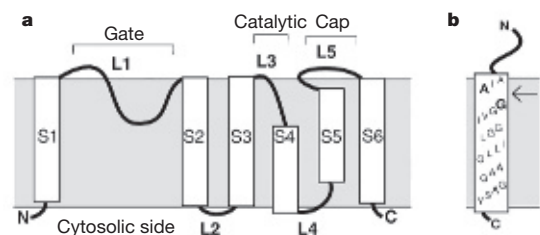


Figure 1 | The membrane topology of a rhomboid protease and its substrate. **a**, The topology of rhomboid protease core domain. **b**, Rhomboid protease cleaves type 1 transmembrane helix near its extracellular end (arrow). For *Toxoplasma gondii* rhomboid substrate MIC2, the membrane-spanning sequence of which is shown here, cleavage occurs at an Ala–Gly (bold) bond three residues below the membrane surface^{45,46}. *Drosophila* Spitz has a similar sequence in this general region, which is recognized and cleaved by Rhomboid-1 (ref. 35).

¹Department of Pharmacology, Yale School of Medicine, 333 Cedar Street, New Haven, Connecticut 06520, USA. †Present address: College of Life Science, Jilin University, Changchun 130023, China.

bilayer (Fig. 2b). The hydrophobic carboxy-terminal region of L1 is positioned within a large V-shaped gap between S1 and S3, and extends out into the lipid phase. This region of L1, as well as the internal cavity, show the highest sequence conservation between rhomboid proteases and, as explained below, are important for the protease function.

Sequence conservation suggests that all rhomboid core structures are likely to be similar, and that the GlpG structure described here is representative of the family (Supplementary Fig. 1). S3, S4 and S6 contain a number of conserved small residues such as glycines at their packing interfaces that render close helical packing possible and also probably provide stabilizing forces²⁸. Unlike GlpG, some rhomboid family members (for example, *Drosophila* Rhomboid-1) appear to contain an additional transmembrane helix attached to the core domain^{12,15}. How this additional segment may affect the structure and function of the core is currently unclear.

GlpG forms a trimer in the crystal (Supplementary Fig. 5). The physiological relevance of the trimer is again unclear. Each protomer contains a complete active site, and is likely to function independently.

Correlation with functional studies

Combining results from this crystallographic study with those from mutagenesis studies on *Drosophila* Rhomboid-1, *Bacillus subtilis* YqgP, *E. coli* GlpG and others^{12,16,21,27}, we identify two regions of GlpG structure to be important for function. The first region is the central cavity that contains all the conserved polar residues of S2, S4 and S6 (His 150, Asn 154, Ser 201, His 254). An extended loop L3, as well as the exposed N terminus of S4, also contribute polar main-chain groups to the cavity. This cavity probably represents the active site of the protease. Substituting each of the conserved residues Gly 199 (L3), Ser 201 (S4) and His 254 (S6) in the cavity invariably abolishes activity^{12,16,21,27} (red in Fig. 3a, b). Mutations of His 150 (S2) and Asn 154 (S2) also affect the activity for some rhomboids, but not all^{12,21,27} (orange in Fig. 3a, b; see below for further discussion).

The hydrophilic cavity is surrounded by transmembrane helices except at its front, where there is a V-shaped gap between S1 and S3 (Fig. 2a). This opening is the likely route by which substrate enters the active site (Fig. 3c): the opening is wide enough to accommodate an α -helix; its lower portion is shallow and hydrophobic, permitting favourable interactions with a hydrophobic substrate while posing no major restrictions on its sequence; its upper portion is connected to the hydrophilic cavity. In the present structure, the lateral opening is blocked by the membrane-embedded loop L1. We postulate that L1 functions as a gate, and may change conformation when substrate binds. Consistent with the possibility that L1 represents another

functionally important region, mutation of Trp 136 or Arg 137—two preferred residues on L1 that are 15 Å away from the active site—either abolishes or reduces protease activity^{12,27} (yellow in Fig. 3a, b). Analysis of the structure of L1, its conservation and the role of the WR motif (tryptophan and arginine) are discussed below.

The active site

The crystal structure supports the hypothesis that rhomboid proteases use a serine–histidine catalytic dyad²⁷. The essential residues Ser 201 and His 254 of GlpG interact via a strong hydrogen bond²⁹

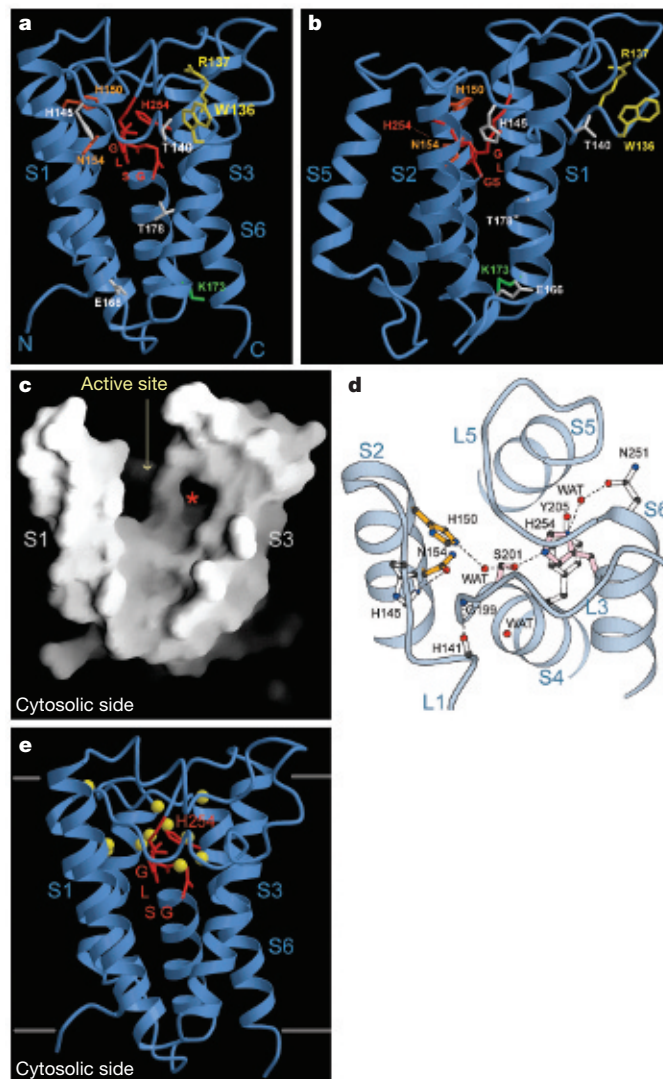


Figure 3 | The membrane-embedded active site of rhomboid protease.

a, b, Mutagenesis studies on *Drosophila* Rhomboid-1 and other rhomboid proteases mapped onto GlpG (viewing angle as in Fig. 2a, b). The GLSG (Gly 199, Ser 201 and His 254) motif and His 254 are shown in red; His 150 and Asn 154 in orange; Trp 136 and Arg 137 in yellow. Alanine substitution of an arginine in Rhomboid-1 that corresponds to Lys 173 of GlpG (green) also affects activity¹²: this residue may have a structural function. Alanine substitutions of those residues shown in white do not affect activity¹². **c**, The internal hydrophilic cavity and its front opening. For clarity, the lateral gate (L1) and the cap (L5) have been omitted. The asterisk marks the location of His 254 towards the back of the cavity. The molecular surface is generated by GRASP⁴⁹. **d**, A detailed view of the active site and the hydrogen bond network surrounding the catalytic Ser 201 and His 254 (pink). WAT, water molecule (isolated red dots). Only parts of L1, S2, L3, S4, S5 and S6 are shown for clarity. **e**, Bound water molecules (yellow spheres) within the active site in relation to the conserved GLSG motif and His 254 (red). The boundaries for the hydrophobic region of the membrane are marked by horizontal lines.

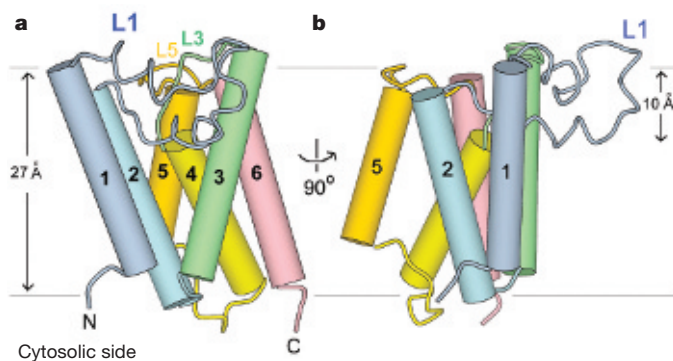


Figure 2 | The overall structure of GlpG. **a**, The front view of a monomer. The transmembrane segments are sequentially labelled 1–6. The two horizontal lines mark the hypothetical boundaries of the hydrocarbon region of a lipid bilayer. **b**, The side view related to **a** by a 90° rotation, as shown. These illustrations, as well as those in Figs 3a, b, d, e and 4, were generated by MOLSCRIPT^{47,48}.

(Fig. 3d). We postulate that the lone base (His 254), like that in several other serine hydrolases^{30–33}, is sufficient to activate the serine for direct nucleophilic attack on substrate³⁴. Although consistent with data showing that rhomboid proteases can be inhibited by 3,4-dichloroisocoumarin^{12,26,27}, a class-specific inhibitor for serine protease, the proposed mechanism lacks direct proof at this time, and therefore remains hypothetical. Asn 154, positioned 8 Å away on the other side of Ser 201, is not within bonding distance to His 254. This latter feature formally eliminates the catalytic triad model¹², explaining why the asparagine is not essential for catalysis, and its requirement can be affected by the context of the reaction^{12,16,21,27}.

A couple of additional interactions may assist His 254 in activating Ser 201 (Fig. 3d). A water molecule mediates hydrogen bonding between His 254 and Asn 251, a frequently observed residue at this position. The ring of His 254 stacks onto that of Tyr 205 (π – π interaction). Most rhomboid proteases contain a tyrosine or phenylalanine at position 205, suggesting that this interaction may be important for the function of the dyad.

The backbone amide of Gly 199, previously thought to contribute to oxyanion binding^{12,34}, is hydrogen bonded to a backbone carbonyl on L1, and points away from the dyad (Fig. 3d). Also, the short loop L5 tightly caps the active site from above. These features suggest that certain structural rearrangements are required to create a functional active site, an idea also raised by a previous study²⁶.

The active site of GlpG contains a number of water molecules (Fig. 3e). The wide distribution of these water molecules, as well as the extensiveness of the protein surface that they interact with, combined with the fact that the capping L5 may open further upon substrate binding, indicate that the reactant water may enter the membrane-embedded active site by different routes from the outside aqueous solution, instead of through a single path or channel.

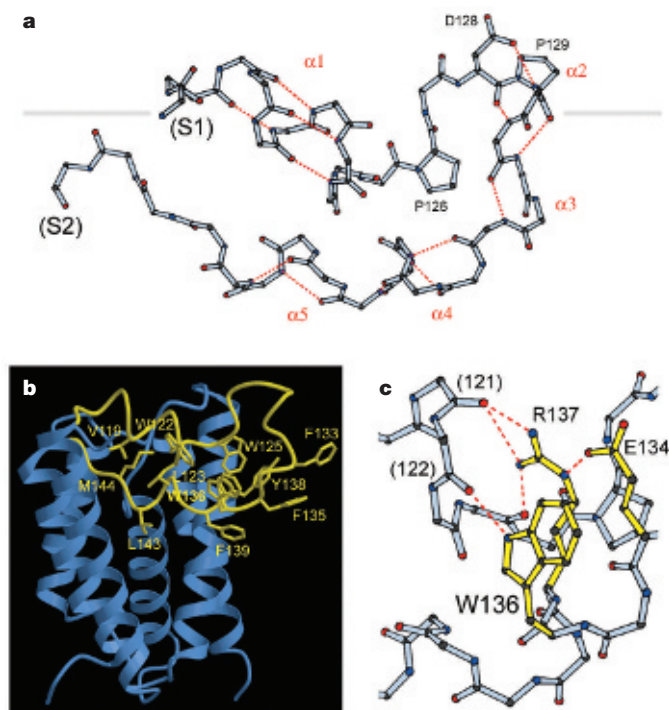


Figure 4 | The structure of the lateral gate. **a**, The main-chain conformation of L1. The approximate membrane boundary is marked by the horizontal line. **b**, Hydrophobic side chains on the membrane-facing side of L1. **c**, A detailed view of the conserved residues Trp 136 and Arg 137 (yellow), and the interactions that they participate in.

The lateral gate

L1 is cradled between S1 and S3, and blocks the lateral opening of the active site (Figs 2a and 3c). It consists mainly of a regular α -helix ($\alpha 1$) and four consecutive 3_{10} helices ($\alpha 2$ – $\alpha 5$) (Fig. 4a). In addition to Trp 136 and Arg 137, there are clear preferences for hydrophobic residues at positions 138, 139 and 143, all facing outwards and interacting with lipid (Fig. 4b). The conserved residue His 145 forms a hydrogen bond with Asn 154 (Fig. 3d). The conservation of these interactions suggests that all rhomboid proteases may use a common structural motif for lateral gating.

L1 has an amphiphilic characteristic: its upper and inward-facing surfaces are polar whereas its lower surfaces are hydrophobic (Fig. 4b). This characteristic is consistent with its peripheral membrane location and its critical position between the active site and lipid. The membrane-embedded string of 3_{10} helices is unusual. Here, the polypeptide contains several turns and a number of polar groups are exposed to lipid, which is energetically unfavourable. This feature may be indicative of a metastable and dynamic structure, consistent with its gating function.

The WR motif has a structural role exclusively within L1 at the membrane–water interface (Fig. 4c). Both residues (tryptophan and arginine) emerge from membrane-embedded locations to interact with the upper structure of L1. Owing to their location towards the tip of the gate away from the main body of the protease (yellow in Fig. 3b), it seems possible that this motif is primarily selected to modulate gate dynamics at the membrane surface. This function would be consistent with previous data showing that the motif is important for activity¹², but is not part of the core catalytic mechanism²⁷, and that its requirement depends on whether the protease is in its native membrane environment²⁷. It was noted that another unrelated membrane protein family also contains a WR motif²⁷.

Mechanistic implications

On the basis of the crystal structure described in this report, we propose a model for rhomboid-mediated intramembrane proteolysis, which is illustrated in Fig. 5. In this model, substrate laterally docks into the gap between two transmembrane helices (S1, S3) of the protease, substituting the lateral gate previously bound there. The surface features of the gap suggest that only the top portion of substrate helix unfolds in the hydrophilic cavity of the protease, and becomes subsequently cleaved (Fig. 3c). This model is consistent with a previous study showing that *Drosophila* Rhomboid-1 recognizes substrate by its top region³⁵. The requirement for substrate to unfold fits not only the need of the nucleophilic reaction, but also the general shape of the protease active site and the preference for helix-destabilizing residues near the top of the substrate³⁵.

Discussion

A fundamental question concerning the mechanism of intramembrane proteolysis is how substrate moves across the phase barrier that separates membrane from water in order to reach the active site of the enzyme. The crystal structure of GlpG illustrates the physical

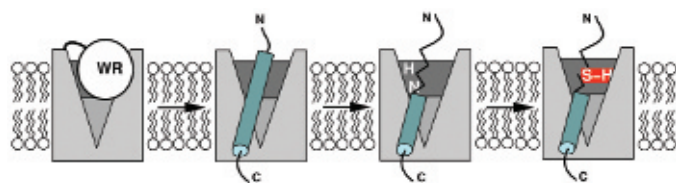


Figure 5 | A possible mechanism of rhomboid-catalysed intramembrane proteolysis. The white circle represents the lateral gate (L1) with the conserved WR motif. The hydrophilic internal cavity is represented by the dark grey area. The transmembrane helix of the substrate is shown in blue. Four residues are highlighted in the active site of the rhomboid protease: the serine–histidine dyad (red), and the conserved histidine and asparagine on S2 (white).

Table 1 | X-ray refinement statistics

Resolution (Å)	40.0–2.1
$R_{\text{work}}/R_{\text{free}}$	0.236/0.252
Number of atoms	
Protein	1,452
Detergent	191
Water	31
B-factors	
Protein	33.6
Detergent	66.1
Water	35.4
R.m.s deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.26

$R_{\text{work}} = \sum |F_o - F_c| / \sum F_o$, R_{free} is the cross-validation R -factor for the test set of reflections (10% of the total) omitted in model refinement. R.m.s., root mean square.

principles underlying one protein machinery that has evolved to facilitate this process. The active site of GlpG is positioned into the membrane at a distance roughly matching the cleavage site of the substrate, but remains separated from the lipid phase by protein structures that surround the active site. Substrate enters and unfolds in the active site through a protein opening that is gated by a specialized mobile structure. This lateral gate has an amphiphilic property that enables it to function smoothly at the membrane–water interface without causing denaturation or aggregation of the enzyme and its transmembrane substrate.

S2P and the presenilins may use different structural motifs to accomplish intramembrane proteolysis (for example, see refs 36, 37). Nevertheless, if lateral gating is a general feature, we note that presenilins do have a highly conserved and hydrophobic sequence—between two transmembrane segments that harbour the catalytic aspartates^{24,38,39}—that could be well positioned to fulfil this function. Many familial Alzheimer's disease mutations are found within this sequence.

METHODS

Crystallization. Crystals were grown at room temperature by the hanging-drop vapour diffusion method from a 5 mg ml^{−1} membrane protein solution in 10 mM Tris-HCl (pH 7.6) and 20 mM nonylglucoside, over a reservoir solution of 3 M NaCl and 100 mM Bis-Tris propane (pH 7.0). Cryo-protection was achieved by stepwise transferral of the crystal to artificial mother liquor containing 25% glycerol. The crystal was flash-frozen in liquid nitrogen. Crystals of the selenomethionine-substituted protein were grown under similar conditions, except that the protein solution also contained 2 mM dithiothreitol and 0.1 mM EDTA. The selenomethionine-substituted crystals took about 1 month to grow to a full size of about 50 μm.

Structure determination. Numerous data sets were collected at 100 K from crystals of the native membrane protein, the selenomethionine-substituted protein, as well as those soaked with various heavy-atom salts, on beamlines X6A, X26C and X29 at Brookhaven National Laboratory–National Synchrotron Light Source (BNL–NSLS). The selenomethionine crystals were small and sensitive to radiation. Diffraction data collected from five crystals were merged to generate the final MAD data set used for solving the selenium substructure and for phasing. All diffraction images were processed by HKL2000 (ref. 40). The ten selenium sites in the asymmetric unit of the crystal were found by hkl2map⁴¹. These sites were used to calculate phases to 2.8 Å resolution. The experimental phases were improved and extended to 2.1 Å resolution by solvent flattening and histogram matching using the program dm of the CCP4 suite⁴², and at a later stage by incorporating phase values calculated from a refined model. The initial polypeptide model was built into the density-modified electron density map using O⁴³, guided by the known selenium coordinates. This model was improved by iterative cycles of manual fitting using O and refinement by CNS⁴⁴. On the basis of clear non-protein densities, detergent and water molecules were added to the model in the final steps of the refinement.

Received 19 July; accepted 18 September 2006.

Published online 11 October 2006.

1. Brown, M. S. & Goldstein, J. L. The SREBP pathway: regulation of cholesterol metabolism by proteolysis of a membrane-bound transcription factor. *Cell* **89**, 331–340 (1997).
2. Mattson, M. P. Pathways towards and away from Alzheimer's disease. *Nature* **430**, 631–639 (2004).

3. Sakai, J. *et al.* Sterol-regulated release of SREBP-2 from cell membranes requires two sequential cleavages, one within a transmembrane segment. *Cell* **85**, 1037–1046 (1996).
4. Rawson, R. B. *et al.* Complementation cloning of S2P, a gene encoding a putative metalloprotease required for intramembrane cleavage of SREBPs. *Mol. Cell* **1**, 47–57 (1997).
5. Haass, C. Take five—BACE and the γ -secretase quartet conduct Alzheimer's amyloid β -peptide generation. *EMBO J.* **23**, 483–488 (2004).
6. Levitan, D. & Greenwald, I. Facilitation of *lin-12*-mediated signalling by *sel-12*, a *Caenorhabditis elegans* S182 Alzheimer's disease gene. *Nature* **377**, 351–354 (1995).
7. De Strooper, B. *et al.* A presenilin-1-dependent γ -secretase-like protease mediates release of Notch intracellular domain. *Nature* **398**, 518–522 (1999).
8. Rudner, D. Z., Fawcett, P. & Losick, R. A family of membrane-embedded metalloproteases involved in regulated proteolysis of membrane-associated transcription factors. *Proc. Natl Acad. Sci. USA* **96**, 14765–14770 (1999).
9. Weihofen, A., Binns, K., Lemberg, M. K., Ashman, K. & Martoglio, B. Identification of signal peptide peptidase, a presenilin-type aspartic protease. *Science* **296**, 2215–2218 (2002).
10. Fluhrer, R. *et al.* A γ -secretase-like intramembrane cleavage of TNF α by the GxGD aspartyl protease SPPL2b. *Nature Cell Biol.* **8**, 894–896 (2006).
11. Friedmann, E. *et al.* SPPL2a and SPPL2b promote intramembrane proteolysis of TNF α in activated dendritic cells to trigger IL-12 production. *Nature Cell Biol.* **8**, 843–848 (2006).
12. Urban, S., Lee, J. R. & Freeman, M. *Drosophila* rhomboid-1 defines a family of putative intramembrane serine proteases. *Cell* **107**, 173–182 (2001).
13. Gallio, M., Sturgill, G., Rather, P. & Kylsten, P. A conserved mechanism for extracellular signalling in eukaryotes and prokaryotes. *Proc. Natl Acad. Sci. USA* **99**, 12208–12213 (2002).
14. Wasserman, J. D., Urban, S. & Freeman, M. A family of rhomboid-like genes: *Drosophila* rhomboid-1 and roughoid/rhomboid-3 cooperate to activate EGF receptor signaling. *Genes Dev.* **14**, 1651–1663 (2000).
15. Koonin, E. V. *et al.* The rhomboids: a nearly ubiquitous family of intramembrane serine proteases that probably evolved by multiple ancient horizontal gene transfers. *Genome Biol.* **4**, R19 (2003).
16. Urban, S., Schlieper, D. & Freeman, M. Conservation of intramembrane proteolytic activity and substrate specificity in prokaryotic and eukaryotic rhomboids. *Curr. Biol.* **12**, 1507–1512 (2002).
17. Sturtevant, M. A., Roark, M. & Bier, E. The *Drosophila* rhomboid gene mediates the localized formation of wing veins and interacts genetically with components of the EGF-R signaling pathway. *Genes Dev.* **7**, 961–973 (1993).
18. McQuibban, G. A., Saurya, S. & Freeman, M. Mitochondrial membrane remodelling regulated by a conserved rhomboid protease. *Nature* **423**, 537–541 (2003).
19. Cipolat, S. *et al.* Mitochondrial rhomboid PARL regulates cytochrome c release during apoptosis via OPA1-dependent cristae remodeling. *Cell* **126**, 163–175 (2006).
20. Brossier, F., Jewett, T. J., Sibley, L. D. & Urban, S. A spatially localized rhomboid protease cleaves cell surface adhesins essential for invasion by *Toxoplasma*. *Proc. Natl Acad. Sci. USA* **102**, 4146–4151 (2005).
21. Maegawa, S., Ito, K. & Akiyama, Y. Proteolytic action of GlpG, a rhomboid protease in the *Escherichia coli* cytoplasmic membrane. *Biochemistry* **44**, 13543–13552 (2005).
22. Brown, M. S., Ye, J., Rawson, R. B. & Goldstein, J. L. Regulated intramembrane proteolysis: a control mechanism conserved from bacteria to humans. *Cell* **100**, 391–398 (2000).
23. Wolfe, M. S. & Kopan, R. Intramembrane proteolysis: theme and variations. *Science* **305**, 1119–1123 (2004).
24. Wolfe, M. S. *et al.* Two transmembrane aspartates in presenilin-1 required for presenilin endoproteolysis and γ -secretase activity. *Nature* **398**, 513–517 (1999).
25. Daley, D. O. *et al.* Global topology analysis of the *Escherichia coli* inner membrane proteome. *Science* **308**, 1321–1323 (2005).
26. Urban, S. & Wolfe, M. S. Reconstitution of intramembrane proteolysis *in vitro* reveals that pure rhomboid is sufficient for catalysis and specificity. *Proc. Natl Acad. Sci. USA* **102**, 1883–1888 (2005).
27. Lemberg, M. K. *et al.* Mechanism of intramembrane proteolysis investigated with purified rhomboid proteases. *EMBO J.* **24**, 464–472 (2005).
28. Russ, W. P. & Engelman, D. M. The GxxxG motif: a framework for transmembrane helix-helix association. *J. Mol. Biol.* **296**, 911–919 (2000).
29. Cleland, W. W., Frey, P. A. & Gerlt, J. A. The low barrier hydrogen bond in enzymatic catalysis. *J. Biol. Chem.* **273**, 25529–25532 (1998).
30. Wei, Y. *et al.* A novel variant of the catalytic triad in the *Streptomyces scabies* esterase. *Nature Struct. Biol.* **2**, 218–223 (1995).
31. Zhou, G. W., Guo, J., Huang, W., Fletterick, R. J. & Scanlan, T. S. Crystal structure of a catalytic antibody with a serine protease active site. *Science* **265**, 1059–1064 (1994).
32. Paetzel, M., Dalbey, R. E. & Strynadka, N. C. Crystal structure of a bacterial signal peptidase in complex with a β -lactam inhibitor. *Nature* **396**, 186–190 (1998).
33. Tjalsma, H. *et al.* Conserved serine and histidine residues are critical for activity of the ER-type signal peptidase SipW of *Bacillus subtilis*. *J. Biol. Chem.* **275**, 25102–25108 (2000).
34. Fersht, A. *Structure and Mechanism in Protein Science: a Guide to Enzyme Catalysis and Protein Folding* (W.H. Freeman, New York, 1999).

35. Urban, S. & Freeman, M. Substrate specificity of rhomboid intramembrane proteases is governed by helix-breaking residues in the substrate transmembrane domain. *Mol. Cell* **11**, 1425–1434 (2003).
36. Ye, J., Dave, U. P., Grishin, N. V., Goldstein, J. L. & Brown, M. S. Asparagine-proline sequence within membrane-spanning segment of SREBP triggers intramembrane cleavage by site-2 protease. *Proc. Natl Acad. Sci. USA* **97**, 5123–5128 (2000).
37. Lazarov, V. K. *et al.* Electron microscopic structure of purified, active γ -secretase reveals an aqueous intramembrane chamber and two pores. *Proc. Natl Acad. Sci. USA* **103**, 6889–6894 (2006).
38. Li, X. & Greenwald, I. Membrane topology of the *C. elegans* SEL-12 presenilin. *Neuron* **17**, 101510–101521 (1996).
39. Li, X. & Greenwald, I. Additional evidence for an eight-transmembrane-domain topology for *Caenorhabditis elegans* and human presenilins. *Proc. Natl Acad. Sci. USA* **95**, 7109–7114 (1998).
40. Otwinowski, Z. & Minor, W. Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
41. Pape, T. & Schneider, T. R. HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *J. Appl. Crystallogr.* **37**, 843–844 (2004).
42. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
43. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
44. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
45. Opitz, C. *et al.* Intramembrane cleavage of microneme proteins at the surface of the apicomplexan parasite *Toxoplasma gondii*. *EMBO J.* **21**, 1577–1585 (2002).
46. Zhou, X. W., Blackman, M. J., Howell, S. A. & Carruthers, V. B. Proteomic analysis of cleavage events reveals a dynamic two-step mechanism for proteolysis of a key parasite adhesive complex. *Mol. Cell. Proteomics* **3**, 565–576 (2004).
47. Kraulis, P. J. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
48. Merritt, E. A. & Murphy, M. E. Raster3D Version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr. D* **50**, 869–873 (1994).
49. Nicholls, A., Sharp, K. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11**, 281–296 (1991).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank V. Stojanoff, H. Robinson, A. Saxena and A. Héroux at BNL NSLS beamlines for help; T. Boggon and J. Schlessinger for sharing the crystallization robot in their laboratories; and B. Turk for sharing the fluorescence spectrometer in his laboratory. X-ray diffraction data for this study were measured at beamlines X6A, X29 and X26C of NSLS. Financial support comes principally from the US Department of Energy, and from the National Institutes of Health. This work was supported by a New Scholar Award in Aging from the Ellison Medical Foundation (to Y.H.) and a gift from the Neuroscience Education and Research Foundation (to Y.H.).

Author Contributions Y.W. and Y.H. purified and characterized GlpG in various detergents. Y.W. crystallized GlpG. Y.H. and Y.W. solved the structure of GlpG and wrote the paper. Y.Z. screened the expression of many constructs and conducted some initial biochemical characterizations.

Author Information The atomic coordinates of GlpG have been deposited in the Protein Data Bank (accession number 2IC8), and will be released upon publication of the paper. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Y.H. (ya.ha@yale.edu).

LETTERS

Lunar activity from recent gas release

Peter H. Schultz¹, Matthew I. Staid² & Carlé M. Pieters¹

Samples of material returned from the Moon have established that widespread lunar volcanism ceased about 3.2 Gyr ago. Crater statistics and degradation models indicate that last-gasp eruptions of thin basalt flows continued until less than 1.0 Gyr ago¹, but the Moon is now considered to be unaffected by internal processes today, other than weak tidally driven moonquakes² and young fault systems³. It is therefore widely assumed that only impact craters have reshaped the lunar landscape over the past billion years. Here we report that patches of the lunar regolith in the Ina structure^{2–5} were recently removed. The preservation state of relief, the number of superimposed small craters, and the ‘freshness’ (spectral maturity) of the regolith together indicate that features within this structure must be as young as 10 Myr, and perhaps are still forming today. We propose that these features result from recent, episodic out-gassing from deep within the Moon. Such out-gassing probably contributed to the radiogenic gases detected during past lunar missions. Future monitoring (including Earth-based observations) should reveal the composition of the gas, yielding important clues to volatiles archived at great depth over the past 4–4.5 Gyr.

The Ina structure (Fig. 1) was first recognized in Apollo images^{4,5}. It was interpreted as a lunar caldera because it occurs on the summit of a 15-km-diameter dome (300 m in relief), is surrounded by a raised but low relief collar, and has a dark halo^{4–7}. Although its fresh appearance indicated a recent activity⁵, age constraints were not set until later^{8,9}. Five criteria constrain the relative ages of lunar features and surfaces: stratigraphy¹⁰; statistics of small craters; retention of photometric properties^{3,9}; preservation of relief and surface texture^{3,11}; and degradation of slopes^{11,12}. Ina contains smooth mounds 5–25 m in relief⁷. It also has numerous smaller mounds and plateaus less than 10 m high surrounded by reflective, low-lying, rough terrains with unresolved relief (<3–6 m high), as illustrated in Fig. 1.

Models of crater slope evolution¹¹ predict that a 100-m-diameter crater on the ejecta of Copernicus (about 1 Gyr old) would have been degraded by subsequent impacts to a slope of less than 1°. For comparison, surface textures on the scale of 5 m in the regolith associated with the 50-Myr-old North Ray crater ejecta at the Apollo 16 landing site have been destroyed, yet they do remain preserved at the 2-Myr-old South Ray crater¹³.

The number density of craters superposed on Ina also indicate an unusually young age (Fig. 2). Only two probable impact craters larger than 30 m can be identified over a count area of 8 km² within the Ina structure. This is comparable to the number of craters on the ejecta of South Ray crater (also near the Apollo 16 landing site), which was dated to ~2 Myr ago. There are a few (~5) additional highly subdued, rimless depressions on top of the mounds that may be degraded impact craters or may be endogenic⁷. The addition of the degraded craters to the crater statistics would increase this age to a maximum of ~10 Myr.

Fresh surfaces within Ina develop on the floors of depressions or along the walls of subdued craters. They are expressed as rubbledd surfaces below a very well defined scarp. The finest scale of this rubble texture, as well as the slopes of the 5–10 m high scarps, cannot be

resolved with the best Apollo imaging (a resolution of 6 m from Apollo panoramic photographs); nevertheless, limits on relief can be set by assumptions of slopes and elevation differences provided by stereo-photogrammetry. It appears that both surfaces and scarps develop at the expense of pre-existing regolith-covered terrains. Therefore, they are not just relicts of ancient features: they must be developing and growing through time.

Ina is not a unique feature. Similar patches of small-scale, low relief patches of rubble occur within the Hyginus central caldera and as extensions of linear rilles in Mare Tranquillitatis³. Apollo photography also reveals very similar features within dark mantling deposits on the edge of Mare Serenitatis¹⁴. In all cases, rubbledd materials comprise the floor of shallow depressions, a location where impact degradation processes (infilling by mass wasting) should be the most efficient. Such fine-scale structure could not date back to the time of mare emplacement, even after allowing for contrasts in

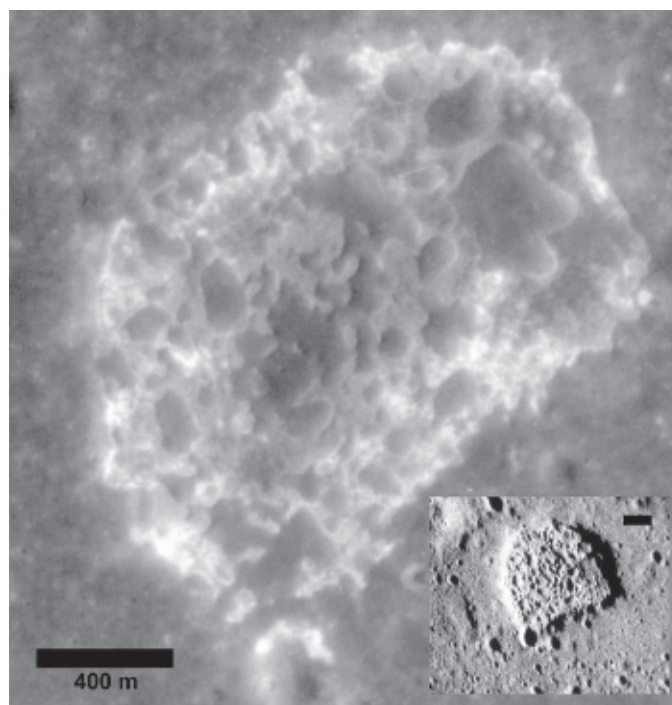


Figure 1 | Apollo photographs of the D-shaped Ina structure. It is 2.8 km in diameter and 60 m in depth. Main panel, the closer view from the Apollo 15 panoramic camera shows the absence of small craters on the interior of the feature and the sharp contacts between smooth mounds of low relief and brighter, hummocky floor. Stereo-photogrammetry demonstrates that the larger smooth mounds can be as high as 25 m, but there are numerous smaller relief features including scarps and floor rubble (bright areas) extending to below the available resolution (<5 m). Inset, a low-illumination view (taken with an Apollo hand-held camera) showing the depression (scale bar, 400 m).

¹Brown University, Geological Sciences, Providence, Rhode Island 02912-1846, USA. ²Planetary Science Institute, 1700 E. Fort Lowell, Suite 106 Tucson, Arizona 85719-2395, USA.

surface strengths (for example, basalts versus regolith), which affect crater scaling and degradation rates¹⁵. Rather, preservation of metre-scale topography requires an age measured in millions, not billions, of years.

Recent data and associated advances in lunar spectroscopy have provided new methods for determining the relative exposure age and composition of the lunar surface^{16–25}. The reflectance properties of the Ina depression and surrounding deposits have been examined using a USGS 100 m per pixel Clementine mosaic¹⁷. Figure 3 provides an overlay of Clementine colour ratio images onto an Apollo 15 panoramic photograph of the Ina structure and the surrounding region. The majority of the soils surrounding Ina have a weak mafic band (shown in purple in Fig. 3), but the 0.8-km-diameter ‘western’ crater exhibits a stronger band (displayed in green) with no change in albedo. The strongest mafic ratios within Ina are directly associated with the interior bright, rubble materials (Fig. 3). The band strength and overall spectral properties of the brightest materials within Ina are most similar to very fresh mare craters in Tranquillitatis, thereby implying fresh exposure of high-titanium basalts.

Plots of spectral ratios against albedo allow examination of the optical maturity and composition of a wide range of lunar materials^{18–24,26,27}. Here, spectral band strength is plotted against the optical albedo (Fig. 4) as a means to assess optical alteration of regional materials due to space weathering²². Band-strength indicates the presence of more mafic and/or less weathered materials within Ina (Fig. 4), consistent with a strewn field of ejecta blocks revealed in high-resolution Apollo panoramic photography of the region.

The trend for the Ina structure (shown in blue in Fig. 4) parallels that for the fresh crater in Lacus Felicitatis (shown in red). As the ultraviolet/visible spectral properties of Ina and the western crater are

also much bluer (higher 415 nm/750 nm ratio) than surrounding deposits, this difference suggests the exposure of an underlying high-titanium mare basalt unit both at Ina and at the western crater. Bright materials within the Ina depression exhibit the strongest mafic band ratio in central Felicitatis, and are interpreted as the least weathered exposures of the underlying high-titanium mare unit—that is, these bright materials have a very young exposure age. For reference, the band strength for materials within Ina approaches that of high-titanium basalts freshly exposed on the wall of Dawes crater²².

Both morphological and spectral criteria indicate that the exposed surfaces within Ina are exceedingly young. In fact, our observations do not preclude the possibility that it is still in the process of formation. The low-lying, fresh exposures within Ina indicate exhumed patches of high-titanium basalt surface beneath a very thick regolith (>12 m) or pyroclastic surface layer, consistent with the regional geologic setting. The original basaltic surfaces date back to at least 3.5 Gyr ago¹⁰, but sudden degassing episodes removed this regolith layer to expose a long-buried basalt surface. The faint halo and raised rim of ejected regolith encircling the Ina structure are consistent with a relatively low energy process with limited ballistic range characterizing the removal process.

Ina is just one of at least four similar endogenic features, most related to a system of west-northwest–east-southeast (WNW–ESE)-trending rilles around the Imbrium impact basin. Ina is at the intersection of northeast–southwest-trending structural elements (radial to Imbrium) and subtle WNW–ESE-trending regional structural elements crossing the Imbrium ejecta. The Hyginus depression is similarly situated at the intersection of the two rilles (Rima Hyginus): a northwest–southeast graben extending to the northwest, and an east-northeast–west-southwest (ENE–WSW) graben to the southwest, which parallels an adjacent en echelon system of rilles (Rima Ariadaeus). Another small irregular depression in Mare Tranquillitatis near the crater Arago also occurs along an extension of the

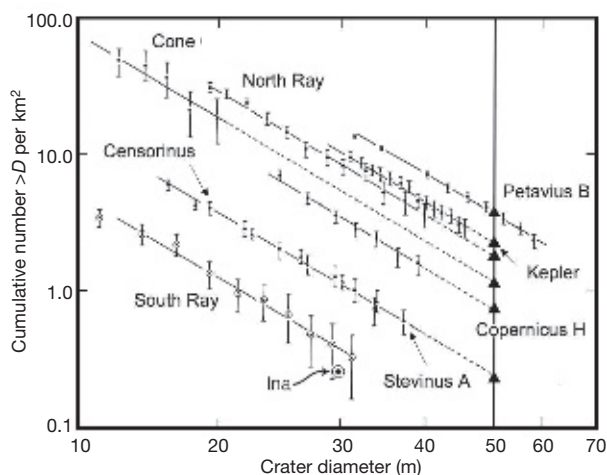


Figure 2 | Comparison between the age of Ina and selected bright-rayed impact craters on the Moon. The age is based on the number density of superimposed small (10–50 m diameter) craters. Plot shows the cumulative number density of craters larger than a given diameter (D), and reflects the relative time since a feature or deposit was formed. Craters sampled by the Apollo mission (North and South Ray, Apollo 16; Cone, Apollo 14) provide approximate calibration for actual time. In an equivalent area, the ejecta deposits around Tycho (109 Myr old) retain more than 8 craters larger than 80 m. Ejecta deposits around North Ray (~50 Myr), Cone (~25 Myr) and South Ray (2 Myr) craters exhibit 80, 30 and 3 craters larger than 30 m, respectively, within the same count area. Only two impact craters larger than 30 m are identified on the interior rubble terrains of Ina. Error bars represent ± 1 standard deviation ($\pm 1\sigma$), based on the total number of craters counted in the selected areas. All crater statistics for the bright-rayed craters were made on the continuous ejecta deposits (more than 0.25 crater radii from the rim crest) or clearly related ejecta deposits (for example, the crater Kepler). Relative ages of the different craters can be more readily compared by extrapolation to a common diameter, which is shown here as the intercepts with a diameter of 50 m (vertical line).

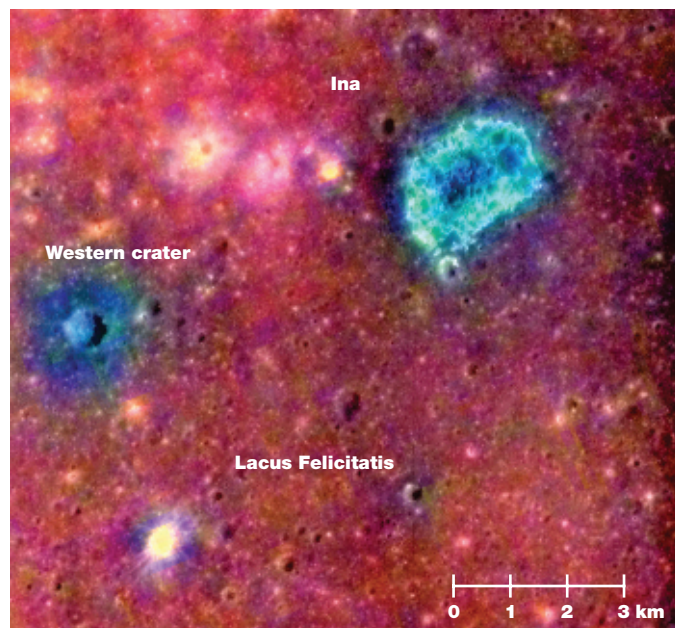


Figure 3 | Clementine colour-ratio composite of the Ina structure and its surrounding region. The false colours indicate variations in composition and age, and correspond to the following ratios: blue, 415 nm/750 nm; green, 750 nm/950 nm; and red, 750 nm/415 nm. Greener regions indicate less mature soils and regions of increased mafic mineralogy. Materials within the Ina depression have distinctly blue ultraviolet/visible ratio values, characteristics that are comparable to high-titanium basalts within Mare Tranquillitatis. The interior of Ina also exhibits a very strong 750 nm/950 nm mafic ratio (displayed in green), relative to surrounding materials, comparable to the smaller, fresh crater to the west. Colour ratio is superimposed on an Apollo 15 panoramic photograph, number 176.

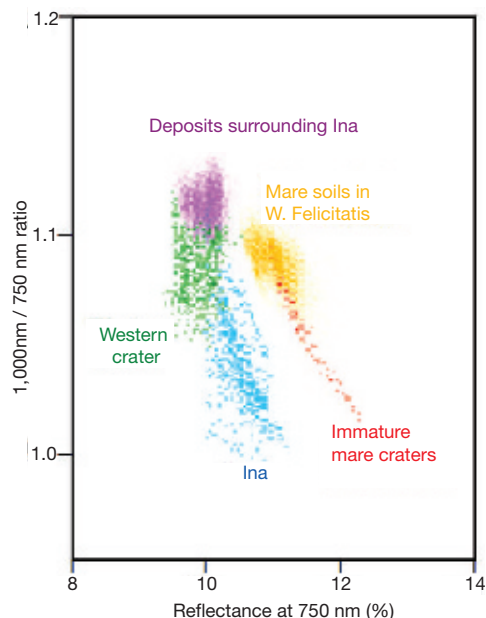


Figure 4 | The ageing of lunar soils revealed by band strength and albedo. Plot shows mafic band strength versus albedo, comparing relative soil maturity (freshness) between the Ina structure and other regions. Mafic band strength (1,000 nm/750 nm ratio) and albedo (750 nm) are plotted for Ina, the surrounding Lacus Felicetatis region, and a separate region of mare deposits in western Felicetatis (see Fig. 3). The 1,000 nm band strength was chosen here because it has been shown to be a reliable indicator of regolith maturity²². Materials within Ina (shown in blue) and the surrounding deposits (purple and green) form a curved low-albedo group, which are distinct from mare deposits to the west (orange and red). Relative band strength for surface materials within Ina is comparable to a 'fresh' (immature) impact crater on the mare and fresher than the nearby impact crater ('western crater') shown in Fig. 3. Consequently, non-impact processes must have modified the interior of Ina recently, consistent with small-scale features (Fig. 1) and crater statistics (Fig. 2).

ENE–WSW (Rima Ariadaeus) system where it intersects a different set of rilles. These occurrences in similar structural settings indicate that volatiles (for example, juvenile CO₂ and even H₂O) trapped deep within the Moon episodically escape along crustal weaknesses, thereby continually freshening the regolith.

Results from the orbiting α -particle spectrometer confirmed the substantive hints from Apollo²⁸ that the Moon is still releasing small amounts of gas from the interior, varying both in space and time²⁹. Such releases currently appear to originate from fresh lunar craters (exposing deeper materials) or in association with pyroclastic deposits²⁹. Ina is adjacent to one of the broad regions having elevated numbers of ²¹⁰Po α -particles, which indicate radon release within the past 60 yr (ref. 29). Consequently, exploring and monitoring Ina and similar structures from both Earth and upcoming lunar missions not only could reveal the nature of the escaping volatiles but also could establish their potential as a resource for future lunar exploration.

Received 27 July; accepted 25 September 2006.

1. Schultz, P. H. & Spudis, P. D. The beginning and end of lunar volcanism. *Nature* **302**, 233–236 (1983).
2. Nakamura, Y. D. et al. New seismic data on the state of the deep lunar interior. *Science* **181**, 49–51 (1973).

3. Schultz, P. H. *Moon Morphology* (Univ. Texas Press, Austin, 1976).
4. Whitaker, E. A. in *Apollo 15 Preliminary Science Report* 25–84–25–85 (NASA SP-289, US Government Printing Office, Washington DC, 1972).
5. El Baz, F. & Worden, A. W. in *Apollo 15 Preliminary Science Report* 25–1–25–25 (NASA SP-289, US Government Printing Office, Washington DC, 1972).
6. El-Baz, F. in *Apollo 17 Preliminary Science Report* 30–13–30–17 (NASA SP-330, US Government Printing Office, Washington DC, 1973).
7. Strain, P. & El Baz, F. The geology and morphology of Ina. *Proc. Lunar Planet. Sci. Conf. XI*, 2437–2446 (1980).
8. Schultz, P. H. in *Workshop on Mare Volcanism and Basalt Petrogenesis* (eds Taylor, L. A. & Longhi, J.) 37–38 (LPI Technical Report 91-03, Lunar and Planetary Institute, Houston, 1991).
9. Schultz, P. H., Staid, M. & Pieters, C. M. Recent lunar activity: evidence and implications. *Lunar Planet. Sci. Conf. XXXI*, 1919 (2000).
10. Wilhelms, D. E. & McCauley, J. F. Geologic map of the near side of the Moon. *US Geol. Surv. Misc. Inv. Map I-703* (1971).
11. Soderblom, L. A. A model for small-impact erosion applied to the lunar surface. *J. Geophys. Res.* **75**, 2655–2661 (1970).
12. Schultz, P. H., Greeley, R. & Gault, D. E. Degradation of small mare surface features. *Proc. Lunar Sci. Conf. VII*, 985–1003 (1976).
13. Wilhelms, D. E. The geologic history of the Moon. *US Geol. Surv. Prof. Pap.* **1348** (1987).
14. Masursky, H. G., Colton, G. W. & El Baz, F. *Apollo over the Moon* (NASA SP-362, US Government Printing Office, Washington DC, 1978).
15. Schultz, P. H., Greeley, R. & Gault, D. E. Interpreting statistics of small lunar craters. *Proc. Lunar Sci. Conf. VIII*, 3539–3564 (1977).
16. Nozette, S. et al. The Clementine Mission to the Moon. *Science* **266**, 1835–1862 (1994).
17. Pieters, C. M., Staid, M. I., Fischer, E. M., Tompkins, S. & He, G. A. Sharper view of impact craters from Clementine data. *Science* **266**, 1844–1848 (1994).
18. Lucey, P. G. et al. Abundance and distribution of iron on the Moon. *Science* **268**, 1150–1153 (1995).
19. Fischer, E. M. & Pieters, C. M. Composition and exposure age of the Apollo 16 Cayley and Descartes regions from Clementine data: Normalizing the optical effects of space weathering. *J. Geophys. Res.* **101** (E1), 2225–2234 (1996).
20. Blewett, D. T., Lucey, P. G. & Hawke, B. R. Clementine images of the lunar sample-return stations: Refinement of FeO and TiO₂ mapping techniques. *J. Geophys. Res.* **102**, 16319–16325 (1997).
21. Lucey, P. G., Blewett, D. T. & Hawke, B. R. Mapping the FeO and TiO₂ content of the lunar surface with multispectral imagery. *J. Geophys. Res.* **103** (E2), 3679–3699 (1998).
22. Staid, M. I. & Pieters, C. M. Integrated spectral analysis of mare craters and soils: Application to eastern nearside basalts. *Icarus* **145**, 122–139 (2000).
23. Lucey, P. G., Blewett, D. T., Taylor, G. J. & Hawke, B. R. Imaging of lunar surface maturity. *J. Geophys. Res.* **105**, 20377–20386 (2000).
24. Le Mouélac, S. et al. Discrimination between maturity and composition of lunar soils from integrated Clementine UV-visible/near-infrared data: Application to the Aristarchus Plateau. *J. Geophys. Res.* **105** (E4), 9445–9455 (2000).
25. Shkuratov, Y. G. et al. Composition of the lunar surface as will be seen from SMART-1: A simulation using Clementine data. *J. Geophys. Res.* **108** (E4), 5020, doi:10.1029/2002JE001971 (2003).
26. Staid, M. I. & Pieters, C. M. Mineralogy of the last lunar basalts: Results from Clementine. *J. Geophys. Res.* **106** (E11), 27887–27900 (2001).
27. Blewett, D. T., Hawke, B. R. & Lucey, P. G. Lunar optical maturity investigations: A possible recent impact crater and a magnetic anomaly. *J. Geophys. Res.* **110**, E04015, doi:10.1029/2004JE002380 (2005).
28. Hodges, R. R. & Hoffman, J. H. Implications of atmospheric ⁴⁰Ar escape on the interior structure of the Moon. *Proc. Lunar Sci. Conf. VI*, 3, 3039–3047 (1975).
29. Lawson, S. L. et al. Recent outgassing from the lunar surface: The Lunar Prospector Alpha Particle Spectrometer. *J. Geophys. Res.* **110**, E09009, doi:10.1029/2005JE002433 (2005).

Acknowledgements We acknowledge S. Posin for his assistance in performing some of the crater statistics.

Author Contributions P.H.S. made the observations, acquired relevant supporting data, and led the writing of the paper; M.I.S. performed the spectral analysis; M.I.S. and C.M.P. contributed to the writing of the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.H.S. (peter_schultz@brown.edu).

Single-mode heat conduction by photons

Matthias Meschke¹, Wiebke Guichard^{1,2} & Jukka P. Pekola¹

The thermal conductance of a single channel is limited by its unique quantum value G_Q , as was shown theoretically¹ in 1983. This result closely resembles the well-known quantization of electrical conductance in ballistic one-dimensional conductors^{2,3}. Interestingly, all particles—irrespective of whether they are bosons or fermions—have the same quantized thermal conductance^{4,5} when they are confined within dimensions that are small compared to their characteristic wavelength. The single-mode heat conduction is particularly relevant in nanostructures. Quantized heat transport through submicrometre dielectric wires by phonons has been observed⁶, and it has been predicted to influence cooling of electrons in metals at very low temperatures due to electromagnetic radiation⁷. Here we report experimental results showing that at low temperatures heat is transferred by photon radiation, when electron–phonon⁸ as well as normal electronic heat conduction is frozen out. We study heat exchange between two small pieces of normal metal, connected to each other only via superconducting leads, which are ideal insulators against conventional thermal conduction. Each superconducting lead is interrupted by a switch of electromagnetic (photon) radiation in the form of a DC-SQUID (a superconducting loop with two Josephson tunnel junctions). We find that the thermal conductance between the two metal islands mediated by photons indeed approaches the expected quantum limit of G_Q at low temperatures. Our observation has practical implications—for example, for the performance and design of ultra-sensitive bolometers (detectors of far-infrared light) and electronic micro-refrigerators⁹, whose operation is largely dependent on weak thermal coupling between the device and its environment.

To get a picture of the radiative thermal coupling, we start by considering two resistors at temperatures T_{e1} and T_{e2} , whose resistances are R_1 and R_2 , respectively, connected via a frequency ($\omega/2\pi$) dependent impedance $Z(\omega)$; see Fig. 1. For simplicity we assume $Z(\omega)$ to be fully reactive, so that only the two resistors emit and absorb noise heating. The net power flow P_v between the two resistors from 1 to 2 due to the electron–photon coupling is then given by^{1,7}:

$$P_v = \int_0^\infty \frac{d\omega}{2\pi} \frac{4R_1R_2}{|Z_t(\omega)|^2} \hbar\omega [n_1(\omega) - n_2(\omega)] \quad (1)$$

Here, $Z_t(\omega) \equiv R_1 + R_2 + Z(\omega)$ is the total series impedance of the circuit, and $n_i(\omega) \equiv [\exp(\hbar\omega/k_B T_{ei}) - 1]^{-1}$ are the boson occupation factors at the temperatures of the resistors $i=1, 2$. Specifically, for a lossless direct connection of the two resistors, $Z(\omega) \equiv 0$, we can integrate equation (1) with the result:

$$P_v^{Z=0} = r_0 \frac{\pi k_B^2}{12\hbar} (T_{e1}^2 - T_{e2}^2) \quad (2)$$

Here $r_0 \equiv 4R_1R_2/(R_1 + R_2)^2$ is the matching factor, which obtains its maximum value of unity when $R_1 = R_2$. Thermal conductance by the photonic coupling, G_v (defined as the linear response of P_v for small

temperature difference $\Delta T \equiv T_{e1} - T_{e2}$ around $T \equiv (T_{e1} + T_{e2})/2$), can then be obtained from equation (2) for the lossless connection as

$$G_v = r_0 G_Q \quad (3)$$

where $G_Q = \frac{\pi k_B^2}{6\hbar} T$ is the quantum of thermal conductance. Thus G_v attains the maximum value for a single transmission channel, G_Q , in a matched circuit. This result is predicted to hold not only for such photon-mediated coupling, but much more generally for carriers of arbitrary exclusion statistics^{10,11} from bosons to fermions^{4,5,12–14}. Note that discussion on massless bosons such as photons or phonons is identical in detail¹, with proper definition of channel transmission in each case. Quantized thermal conductance was observed for the first time in phonons in ref. 6.

Owing to its relatively weak temperature dependence, $\propto T$, electron–photon coupling is expected to become the dominant relaxation means at sufficiently low T . The competing electron–phonon thermal conductance G_{ep} behaves normally as $G_{ep} \approx 5\Sigma\Omega T^4$, where Σ is a material parameter and Ω is the volume of the resistor. This result derives from the expression for heat flux from electrons to lattice, $P_{ep} \approx \Sigma\Omega(T_{ei}^5 - T_0^5)$, where T_{ei} and T_0 are the temperatures of the electrons in the resistor and of the lattice, respectively⁸. Equating G_v from equation (3) and G_{ep} , one finds the cross-over temperature, T_{cr} , below which the photonic conductance should

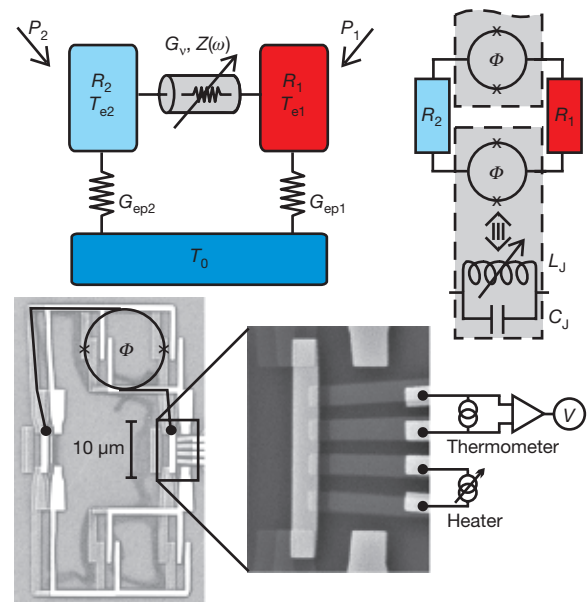


Figure 1 | The system under investigation. At the top, we show thermal (left) and electrical (right) models; at the bottom, a scanning electron micrograph of the device (left), and a magnified view of resistor 1 with four adjoining NIS and two NS contacts (right). See text for details.

¹Low Temperature Laboratory, Helsinki University of Technology, PO Box 3500, 02015 TKK, Finland. ²Université Joseph Fourier and LCM-CNRS, BP 166, 25 Avenue des Martyrs, 38042 Grenoble cedex 9, France.

dominate⁷: $T_{cr} = [r_0 \pi k_B^2 / (30 \hbar \Sigma \Omega)]^{1/3}$. For typical metals, for which $\Sigma \approx 10^9 \text{ W m}^{-3} \text{ K}^{-5}$, for mesoscopic resistors with $\Omega \approx 10^{-20} \text{ m}^3$, and for matching where r_0 is not too low as compared to unity, one obtains $T_{cr} \approx 100\text{--}200 \text{ mK}$. Such temperatures are in the range of experiments that we describe here. By state-of-the-art electron-beam lithography one can obtain metallic islands of volumes as small as $< 10^{-24} \text{ m}^3$ (ref. 15), and there T_{cr} could be as high as several kelvin.

To investigate the electron–photon thermal conduction experimentally, we have embedded a tunable impedance between the two resistors. This allows us to measure the modulation of P_v , or G_v , in response to the externally controllable impedance $Z(\omega)$. In practice the two resistors are connected to each other symmetrically by two aluminium superconducting lines, interrupted by a DC-SQUID (superconducting quantum interference device) in each line, as shown in the electron micrograph and in the electrical model in Fig. 1. These SQUIDs serve as the thermal switches between the resistors, as will be described below.

The structures have been fabricated by electron-beam lithography and three angle shadow evaporation. The film thickness of the superconducting lines is 20 nm. The two AuPd resistors are nominally identical—6.6 μm long, 0.8 μm wide and 15 nm thick. Their resistances are $R_i \approx 200 \Omega$ each. One of them, say R_1 , is connected by four NIS (normal–insulator–superconductor) tunnel junctions to external aluminium leads to allow for thermometry¹⁶ and Joule heating. The normal state resistance of each NIS junction is about 50 k Ω . The resistors are connected by direct NS contacts to the superconducting lines in between, without a tunnel barrier. They are, however, long enough such that they are not noticeably affected by proximity superconductivity. This is verified by the measured tunnel characteristics of the nearby NIS tunnel junctions. Owing to the superconductors being at a low working temperature, which is typically a factor of ten below the critical temperature $T_C \approx 1.2 \text{ K}$ of aluminium, the normal electronic thermal conductance along the lines is efficiently suppressed.

Each DC-SQUID can be modelled as a parallel connection of a Josephson inductance L_J and capacitance C_J . $L_J \approx \hbar / (2eI_C)$ can be tuned by external magnetic flux Φ threading through the DC-SQUID loop, as the critical current is $I_C \approx I_{C0} |\cos(\pi\Phi/\Phi_0)|$. Here I_{C0} is the Ambegaokar–Baratoff critical current¹⁷ determined by geometry and materials, and $\Phi_0 = h/2e \approx 2 \times 10^{-15} \text{ Wb}$ is the flux quantum. Capacitance C_J is a constant determined by geometry and materials.

The length of the lines connecting resistors R_1 and R_2 is $\sim 30 \mu\text{m}$, that is, much shorter than the typical thermal wavelength $\lambda_{th} = 2\pi\hbar c / (k_B T)$, which is several centimetres at 100 mK; here $c \approx 10^8 \text{ m s}^{-1}$ is the speed of light on a silicon substrate. Therefore we do not need to consider a distributed electrical model with a transmission line, but instead $Z(\omega)$ is effectively a lumped series connection of two LC circuits, that is, $Z(\omega) = i2\omega L_J / [1 - (\omega/\omega_0)^2]$. Here $\omega_0 = (L_J C_J)^{-1/2}$, and we have assumed for simplicity that the two DC-SQUIDs are identical and that they are exposed to the same magnetic field. This is expected to be a good approximation in view of the symmetric experimental configuration. The net heat flow between 1 and 2 is then given by

$$P_v = \frac{\pi k_B^2}{12\hbar} (r_1 T_{e1}^2 - r_2 T_{e2}^2) \quad (4)$$

where the matching parameters r_i now depend on temperature as:

$$r_i = \frac{6r_0}{\pi^2} \int_0^\infty dx \frac{x}{e^x - 1} \left[1 + \frac{(\omega_{th,i} \tau_R)^2 x^2}{[1 - (\omega_{th,i}/\omega_0)^2 x^2]^2} \right]^{-1} \quad (5)$$

Above we have defined $\omega_{th,i} \equiv k_B T_{e,i} / \hbar$ and $\tau_R \equiv L_J / [(R_1 + R_2)/2]$.

For the full description, we still need a thermal model incorporating the two competing relaxation mechanisms, due to photons and phonons, respectively. This is schematically depicted in Fig. 1, where

the temperature in each resistor tends to relax via electron–phonon coupling G_{epi} to the constant temperature T_0 of the bath, and towards a common temperature of the two resistors via the tunable photonic conductance G_v . P_i denotes the external heat leak into resistor i : owing to wire connections, P_1 has a significant non-zero value even in the absence of intentional heating, whereas P_2 turns out to be very small, as the corresponding resistor is not connected directly to external leads. We may describe the steady state of the system by power balance equations:

$$P_i = \pm \frac{\pi k_B^2}{12\hbar} (r_1 T_{e1}^2 - r_2 T_{e2}^2) + \Sigma \Omega_i (T_{ei}^5 - T_0^5), i = 1, 2 \quad (6)$$

where $\pm$ equals $+$ for $i = 1$, and $-$ for $i = 2$, and Ω_i is the volume of resistor i . Equation (6) combined with equation (5) can be solved numerically to obtain the temperatures T_{ei} under given conditions. We note that the thermal analysis of the system would be even simpler, if resistor 2 were very large in size (but still with the same resistance) to fix T_{e2} at T_0 . Yet such a design would introduce complications in fabrication, and also the simple lumped element electrical analysis would eventually fail.

The experiments were performed in a ^3He – ^4He dilution refrigerator at temperatures from 30 mK up to several hundred mK. All the measurement wiring was carefully filtered and essentially only DC signals were used. One of the NIS tunnel junction pairs connected to resistor 1 was used as a thermometer by applying a small (2.4 pA) DC current I_p through it and by measuring the corresponding temperature dependent voltage V_p . When biased at a low enough current, high sensitivity is obtained without excessive self-heating or self-cooling effects⁹. The NIS thermometer is then calibrated by measuring V_p against the bath temperature T_0 , by slowly sweeping the temperature of the mixing chamber through the range of interest, from about 40 mK up to about 500 mK. Figure 2a shows such a calibration run. The dependence of $R \equiv V_p/I_p$ on temperature turns out to be linear in the range 100–300 mK, which is as expected at

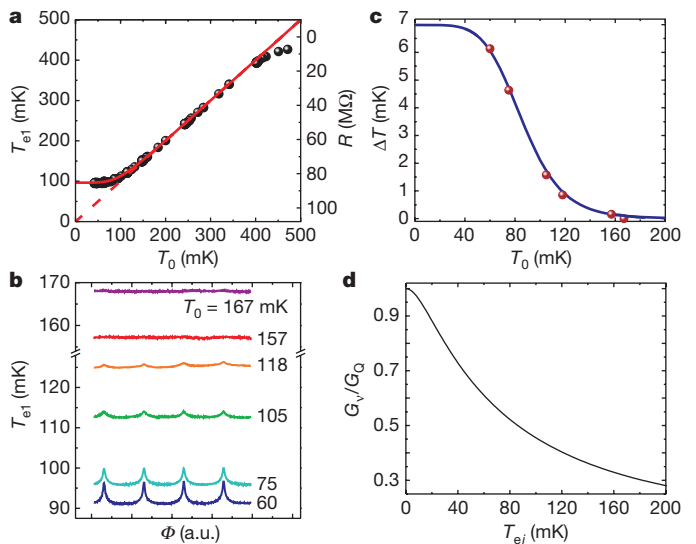


Figure 2 | Results of the measurements in the absence of extra heating.

a, The NIS thermometer calibration. The dots indicate the measured resistance (right scale) at the bias current of 2.4 pA as a function of bath temperature T_0 . The lines show calculated T_{e1} (left scale) versus T_0 for $P_1 = 1 \text{ fW}$ (solid line) and for $P_1 = 0$ (dashed line). **b**, Flux modulation of T_{e1} recorded at the values of T_0 indicated. **c**, The measured amplitude (symbols) of temperature modulation plotted against T_0 and compared to the theoretical model (line) as described in the text, assuming $P_1 = 1 \text{ fW}$ according to **a**. **d**, The ratio of the photonic thermal conductance G_v in our circuit at $\Phi = 0$ and the quantum of thermal conductance G_Q as a function of the electron temperature T_{e1} . This plot is a calculated result based on the experimentally determined values of the system parameters, as given in the text.

temperatures well below T_C (refs 9, 16). At low temperatures, R saturates, which we ascribe predominantly to residual heat input into resistor 1 owing to coupling to the environment and to self-heating. By setting $P_1 = 1$ fW, we obtain the measured R versus T_0 curve through equation (6) at half-integer values of Φ/Φ_0 , where the photonic contribution can be neglected. We use this value of P_1 in analysing the actual measurements.

The measurements of thermal coupling are then done by stabilizing the bath at a desired temperature. Then we apply slow sweeps ($\sim 1 \Phi_0 \text{ min}^{-1}$) of external magnetic flux, which threads the two DC-SQUID loops nominally identically. A typical amplitude of the field sweep is such that it corresponds to 4–5 flux quanta through each DC-SQUID. In our geometry, this corresponds to a field of about 100 μT . The NIS thermometer reading is then averaged over several tens of such field sweeps, to measure accurately the periodic variation of T_{e1} in response to the field sweep.

Figure 2b shows results of a measurement at a few bath temperatures as described above. We see that the modulation of T_{e1} in response to magnetic flux Φ is about 6 mK at the lowest bath temperature and it decreases monotonically when T_0 is increased. Based on this data and our electrical and thermal models above, we expect that the maxima in T_{e1} correspond to the weakest electron–photon coupling at half-integer values of Φ/Φ_0 . In Fig. 2c we plot with circles the corresponding modulation amplitude of T_{e1} between its maximum and minimum values, $\Delta T \equiv T_{e1,\text{max}} - T_{e1,\text{min}}$, as a function of T_0 for the experimental data in Fig. 2b. We have added to the figure a solid line from our theoretical model, assuming $R_1 = R_2$, $P_1 = 1$ fW, $P_2 = 0$ and $I_{C0} = 20$ nA. The last value is a fit parameter, as we cannot measure it directly in our geometry, but it is in line with the measured values of critical currents of DC-SQUIDS that we fabricated with the same parameters separately. We set $C_j = 15$ fF, based on the geometry of the device. Additionally, we use a typical value $\Sigma = 2 \times 10^9 \text{ W K}^{-5} \text{ m}^{-3}$ (ref. 8), and $\Omega_i = 6 \times 10^{-20} \text{ m}^3$ for both resistors. The latter corresponds to the volume of each resistor excluding the overlap areas of NS contacts. The agreement between the experiment and the model is very good, and we therefore use these very realistic parameters in analysing all

the results reported here. These data imply (Fig. 2d) that r_i at integer values of Φ/Φ_0 lies in the range 0.6–0.3, when we vary temperature from ~ 60 mK (the approximate value of T_{e2} at the minimum bath temperature) up to 200 mK, that is, the electron–photon conductance is about one-half of its quantum value in our experiment. Note that G_v is expected to approach the quantum of thermal conductance upon lowering temperature, as the thermal frequency $\omega_{\text{th},i}$ decreases linearly with T_{eib} and the line impedance at low frequencies, determined by L_j , decreases likewise.

Next we present experiments where external power was applied to resistor 1, using the second pair of tunnel junctions connected to it as a heater. Figure 3 demonstrates such measurements at different bath temperatures. Compared to the data at the lowest bath temperature of 60 mK, where an essentially monotonic decrease of ΔT can be observed, the intermediate bath temperature data demonstrates a non-monotonic dependence: there is an initial increase of the signal on increasing the heating, and then slow decrease towards higher power levels and higher T_{e1} . Finally, at the highest values of T_0 , the signal is first absent but emerges upon increasing the input power. This behaviour arises because by applying Joule heating to just one resistor, we can establish a larger temperature difference between the two. Yet at large enough levels of P_1 , the overall temperature of the system increases, and the significance of the photon coupling with respect to electron–phonon coupling is diminishing, and the temperature modulation becomes very weak. All these dependences are fully consistent with our theoretical modelling of the system, and we obtain quantitative agreement with the data by using the same electrical and thermal parameters of the system as when modelling data of Fig. 2. The theoretical lines in Fig. 3 are the result of this modelling. In the inset of Fig. 3 we show a few full modulation curves of T_{e1} versus Φ at a bath temperature of 75 mK. The black lines are from experiment and the red ones from the theoretical model. The theoretical lines capture the main features of the shape of the experimental curves.

Our experiment demonstrates what was proposed recently⁷: it is not justified to assume that an electron system like the one discussed here would be efficiently decoupled from the environment once the electron–phonon coupling is suppressed at low temperatures. The observed $\propto T$ photonic conductance thus has implications for the performance and design of micro-bolometers and calorimeters, where efficient suppression of thermal coupling is usually taken for granted. Owing to better coupling to the environment by photon conductance, the expected sensitivity of such devices is reduced, and their noise is enhanced. On the other hand, this mechanism could possibly provide a way to tune the thermal coupling of a bolometer to the heat bath in order to optimize its operation, which is a trade-off between sensitivity and bandwidth. The radiation of heat could also help to remove excessive heat—for example, from dissipative shunt resistors of ultra-sensitive or very low temperature SQUIDS¹⁸, or at the back side of an electronic micro-refrigerator¹⁹. Furthermore, the photonic coupling could act as a mediator of decoherence: for example, in solid-state quantum coherent devices²⁰. The strength of this harmful effect depends (as in our present experiment) critically on the matching between the noise source and the system that it affects.

Received 16 May; accepted 21 September 2006.

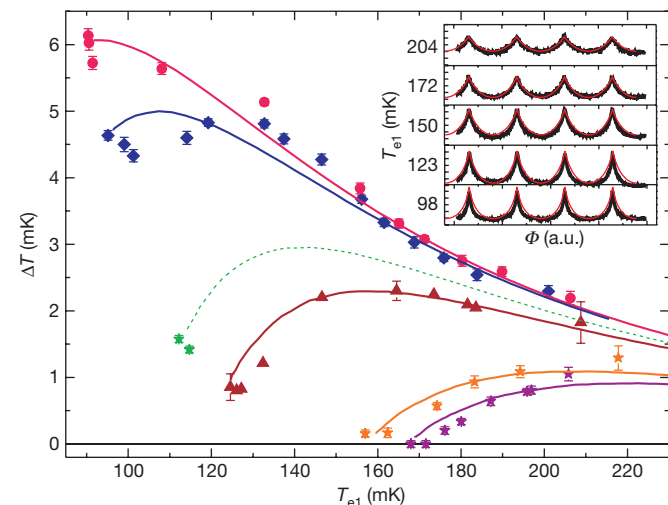


Figure 3 | Results of the measurement of ΔT with variable amounts of Joule heating applied to resistor 1. ΔT is the modulation of T_{e1} when flux is varied; Joule heating was in the range 0–60 fW. Different sets of data and lines correspond to different bath temperatures T_0 . From top to bottom, $T_0 = 60$ mK, 75 mK, 105 mK, 118 mK, 157 mK and 167 mK. The symbols refer to experimental data, and lines are from the theoretical model described. The error bars display the standard deviation between the measured thermometer reading (versus Φ) and the empirical fit function to it. Inset, primary data in form of T_{e1} against flux, under different input power levels and at $T_0 = 75$ mK. Black lines are from the experiment and red ones from the theory. The full temperature range in each panel is 6 mK.

1. Pendry, J. B. Quantum limits to flow of information and entropy. *J. Phys. A* **16**, 2161–2171 (1983).
2. van Wees, B. J. *et al.* Quantized conductance of point contacts in a two-dimensional electron-gas. *Phys. Rev. Lett.* **60**, 848–850 (1988).
3. Wharam, D. A. *et al.* One-dimensional transport and the quantisation of the ballistic resistance. *J. Phys. C* **21**, L209–L214 (1988).
4. Rego, L. G. C. & Kirczenow, G. Fractional exclusion statistics and the universal quantum of thermal conductance: A unifying approach. *Phys. Rev. B* **59**, 13080–13086 (1999).
5. Blencowe, M. P. & Vitelli, V. Universal quantum limits on single-channel information, entropy, and heat flow. *Phys. Rev. B* **62**, 052104 (2000).

6. Schwab, K., Henriksen, E. A., Worlock, J. M. & Roukes, M. L. Measurement of the quantum of thermal conductance. *Nature* **404**, 974–977 (2000).
7. Schmidt, D. R., Schoelkopf, R. J. & Cleland, A. N. Photon-mediated thermal relaxation of electrons in nanostructures. *Phys. Rev. Lett.* **93**, 045901 (2004).
8. Roukes, M. L., Freeman, M. R., Germain, R. S., Richardson, R. C. & Ketchen, M. B. Hot electrons and energy transport in metals at millikelvin temperatures. *Phys. Rev. Lett.* **55**, 422–425 (1985).
9. Giazotto, F., Heikkilä, T. T., Luukanen, A., Savin, A. M. & Pekola, J. P. Opportunities for mesoscopics in thermometry and refrigeration: Physics and applications. *Rev. Mod. Phys.* **78**, 217–274 (2006).
10. Wilczek, F. Quantum mechanics of fractional-spin particles. *Phys. Rev. Lett.* **49**, 957–959 (1982).
11. Haldane, F. D. M. “Fractional statistics” in arbitrary dimensions: A generalization of the Pauli principle. *Phys. Rev. Lett.* **67**, 937–940 (1991).
12. Kane, C. L. & Fisher, M. P. A. Thermal transport in a Luttinger liquid. *Phys. Rev. Lett.* **76**, 3192–3195 (1996).
13. Kane, C. L. & Fisher, M. P. A. Quantized thermal transport in the fractional quantum Hall effect. *Phys. Rev. B* **55**, 15832–15837 (1997).
14. Greiner, A., Reggiani, L., Kuhn, T. & Varani, L. Thermal conductivity and Lorenz number for one-dimensional ballistic transport. *Phys. Rev. Lett.* **78**, 1114–1117 (1997).
15. Pashkin, Yu A., Nakamura, Y. & Tsai, J. S. Room-temperature Al single-electron transistor made by electron-beam lithography. *Appl. Phys. Lett.* **76**, 2256–2258 (2000).
16. Nahum, M. & Martinis, J. M. Ultrasensitive-hot-electron microbolometer. *Appl. Phys. Lett.* **63**, 3075–3077 (1993).
17. Ambegaokar, V. & Baratoff, A. Tunneling between superconductors. *Phys. Rev. Lett.* **10**, 486–489 (1963).
18. Savin, A. M., Pekola, J. P., Averin, D. V. & Semenov, V. K. Thermal budget of superconducting digital circuits at subkelvin temperatures. *J. Appl. Phys.* **99**, 084501 (2006).
19. Clark, A. M., Williams, A., Ruggiero, S. T., van den Berg, M. L. & Ullom, J. N. Practical electron-tunneling refrigerator. *Appl. Phys. Lett.* **84**, 625–627 (2004).
20. Makhlin, Yu, Schön, G. & Shnirman, A. Quantum-state engineering with Josephson-junction devices. *Rev. Mod. Phys.* **73**, 357–400 (2001).

Acknowledgements We thank M. Paalanen, D. Averin, A. Luukanen, H. Pothier, F. Hekking and G. Schön for comments, and the Academy of Finland (TULE) and the EC-funded ULTI project for financial support.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.P.P. (jukka.pekola@tkk.fi).

Boundary lubrication under water

Wuge H. Briscoe¹, Simon Titmuss¹, Fredrik Tiberg^{1,†}, Robert K. Thomas¹, Duncan J. McGillivray^{1,†} & Jacob Klein^{1,2}

Boundary lubrication, in which the rubbing surfaces are coated with molecular monolayers, has been studied extensively for over half a century^{1–7}. Such monolayers generally consist of amphiphilic surfactants anchored by their polar headgroups; sliding occurs at the interface between the layers, greatly reducing friction and especially wear of the underlying substrates. This process, widespread in engineering applications, is also predicted to occur in biological lubrication via phospholipid films^{8,9}, though few systematic studies on friction between surfactant layers in aqueous environments have been carried out^{5,10}. Here we show that the frictional stress between two sliding surfaces bearing surfactant monolayers may decrease, when immersed in water, to as little as one per cent or less of its value in air (or oil). We attribute this to the shift of the slip plane from between the surfactant layers, to the surfactant/substrate interface. The low friction would then be due to the fluid hydration layers surrounding the polar head groups attached to the substrate. These results may have implications for future technological and biomedical applications.

Normal and shear forces $F_{\perp}$ and F_s respectively are measured between two atomically smooth curved mica surfaces (radius of curvature $R \approx 1$ cm) using a surface force balance described previously¹¹. This is carried out first in dry air (in the presence of P_2O_5), then when bearing the double-chained cationic surfactant DDunAB¹², N,N -dimethyl- N,N -diundecylammonium bromide with structure $(CH_3(CH_2)_{10})_2N^+(CH_3)_2Br^-$, typical of a large class of similar amphiphilic surfactants, both in dry air and in (surfactant-free) water. We carried out some ten separate experiments on the DDunAB with surfactant layers formed by incubation times ranging from 10 min to 120 min.

Normalized force $F_{\perp}$ versus separation D profiles, $F_{\perp}(D)/R$, including controls between surfaces of bare mica before surfactant self-assembly, are shown in Fig. 1. The surfactant-coated surfaces experience the expected strong adhesion both in dry air and in water^{13,14}. The total surfactant layer thickness in dry air contact is $D_0 = 2.7 \pm 0.4$ nm, smaller than twice the total extended molecular length of the surfactant, 3.8 nm, indicating that the molecules adsorb onto each mica surface with the hydrocarbon tails tilted. On adding water the layers swelled slightly, as revealed by an increase of the contact separation after the first approach and adhesion in water (averaged over all experiments), by an amount $\Delta D_0 = 0.51 \pm 0.4$ nm. The magnified-scale inset in Fig. 1 reveals the long-range attraction from $D \approx 30$ nm, before the jump into contact characteristic of hydrophobic interactions¹⁴. The strong adhesion that the surfaces experience in contact (in both air and water), deforms them elastically¹⁵ to form a flattened, roughly circular contact zone between them with an area A (see interference fringes in Fig. 2) of the order of $1,000 \mu m^2$. It is across this flattened, strongly adhering region that we measure the friction force F_s between the surfaces.

The frictional response between the surfactant-coated surfaces as the top surface moves laterally back-and-forth after a first approach is

shown in Fig. 2. In dry air (trace B in Fig. 2) the surfaces remain rigidly coupled to each other for all applied shear amplitudes Δx_0 (10–1,800 nm). The friction force F_s between the surfaces must thus exceed the maximum applied shear force ($F_s > F_{\max} = 200$ – $250 \mu N$), corresponding to a sliding shear stress between the surfaces, $\sigma_s = F_s/A > 0.2$ MPa. This is consistent with the range of previous

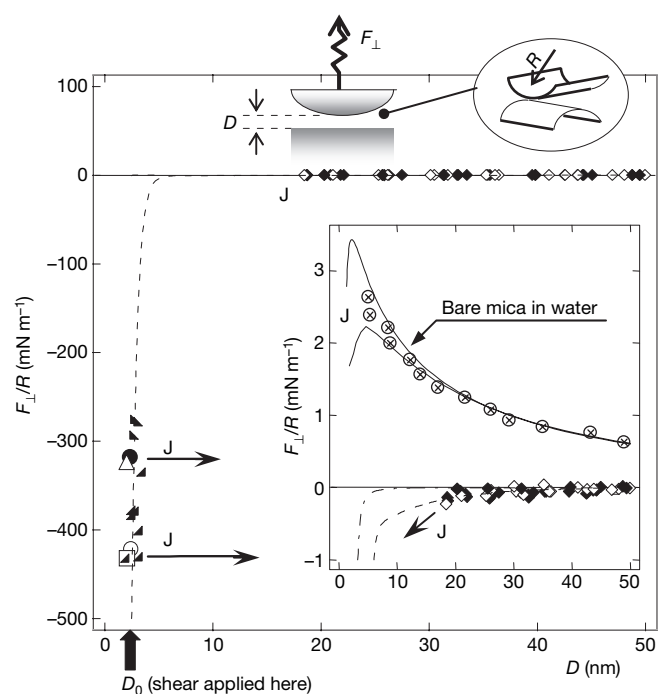


Figure 1 | Normal interactions $F_{\perp}/R$ between DDunAB-coated crossed-cylindrical mica surfaces (radius of curvature R is ~ 1 cm), as a function of closest separation D . Forces $F_{\perp}$ were measured with a surface force balance¹¹ (cartoon). Different symbols indicate separate experiments (10-min or 120-min layers, see Methods). The open triangle and closed circle represent adhesive contact in dry air (mean of several measurements). The closed and open diamonds represent interaction profiles across water with no added salt, up to jumps (J) into adhesive contact. The open square and circle indicate adhesive contact following first approach in water. The adhesion in water on subsequent approaches following shear is similar or slightly decreased ($\blacktriangle$ is 120 min, and $\blacksquare$ is 10 min). The inset shows interactions on an enlarged scale: crossed circles represent interactions between the bare mica surfaces (solid curves show expected DLVO double-layer forces¹³. Upper solid curve, constant surface charge density of 0.0048 C m^{-2} ; lower solid curve, constant surface potential of 155 mV; both with $1.5 \times 10^{-5} \text{ M}$ 1:1 background electrolyte concentration). Other symbols as in the main figure. The lowest curves are van der Waals attraction (dot-dashed curve) and an empirical double exponential function (dashed curve) characteristic of hydrophobic attraction¹⁴. All measurements are at $25 \pm 0.5^\circ \text{C}$.

¹Physical and Theoretical Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QZ, UK. ²Weizmann Institute of Science, Rehovot 76100, Israel. [†]Present addresses: Camrus AB, Ideon Science Park, Sölvegatan 41, SE-223 70 Lund, Sweden (F.T.); Department of Biophysics, Johns Hopkins University, Baltimore, Maryland 21218, USA (D.J.M.).

studies^{2,10,16} of the frictional shear stress between surfactant-coated surfaces in dry air at comparable normal stresses and velocities. Once water is introduced between the surfactant-coated surfaces, the sliding frictional stress between them when in adhesive contact is strikingly reduced relative to air, despite the stronger adhesion under water (Fig. 1). A typical friction trace C following initial contact is shown in Fig. 2.

Figure 3 summarizes the measured sliding friction stresses $\sigma_s = F_s/A$ between the surfactant layers under water at different shear velocities v_s , revealing a weak velocity dependence in line with earlier studies^{2,6,10}. The shaded part of the inset to Fig. 3 shows the range of frictional stresses taken from earlier studies^{2,3,10,16}, covering a wide spectrum of boundary surfactants and conditions in air. The most remarkable feature of our data is the very substantial reduction—by up to 99% or more—of σ_s under water relative to its value in air.

What is the origin of this very large decrease in the friction when the sliding occurs under water? In the case of friction between similar surfactant layers in air, a coherent picture has emerged (for example, see refs 3 and 17), correlating the frictional stress σ_s during sliding with the loading/unloading adhesion hysteresis $\Delta\gamma$ between the boundary layers. The relation $\Delta\gamma/\delta \approx \sigma_s$, where δ is a microscopic length scale (of the order of 1 nm or less), accounts well for the frictional dissipation in a wide range of conditions (in air) as the surfactant-coated surfaces slide past each other³. However, we attribute the striking reduction in friction under water relative to dry air observed here to a conceptually quite different origin. It is due, we believe, to the shift of the slip plane from between the boundary lubricant layers to the surfactant–substrate interface.

This occurs because of the facile penetration of water and consequent hydration of the anchoring polar headgroups at the solid substrate^{3,17,18}, consistent with the known high mobility of water in comparable surfactant layers^{19,20}. It is significant in this context that the swelling $\Delta D_0/2$ of each monolayer on adding water

(0.26 ± 0.2 nm) is close to the swelling due specifically to hydration of headgroups (0.25 nm), measured¹⁸ for similar confined surfactant monolayers on going from the dry state to immersion in water. The resulting hydration sheaths efficiently lubricate the lateral motion of the headgroups on the surface, in strong analogy with the recently observed lubrication provided by hydrated metal ions between solid surfaces^{21,22}. As a result, the interface with least resistance to shear becomes that between the surfactant headgroups and the substrate, rather than the surfactant–surfactant interface (where sliding classically occurs in air or oil). It is therefore to that weaker interface that the slip plane reverts as the surfaces slide past each other.

To test this hypothesis—which contrasts with the accepted mechanism for boundary lubrication in air or oil—more directly, we carried out the following three experiments. First, we determined the adhesion hysteresis $\Delta\gamma$ between the surfactant layers both in dry air and under water, by monitoring the contact area during loading/unloading cycles (Methods section). The results, shown in Fig. 4, reveal that the magnitude of $\Delta\gamma$ is in the range (20 ± 2) mJ m⁻² both in air and under water. Putting $\Delta\gamma/\delta \approx \sigma_s$, with $\delta = 1$ nm, yields a

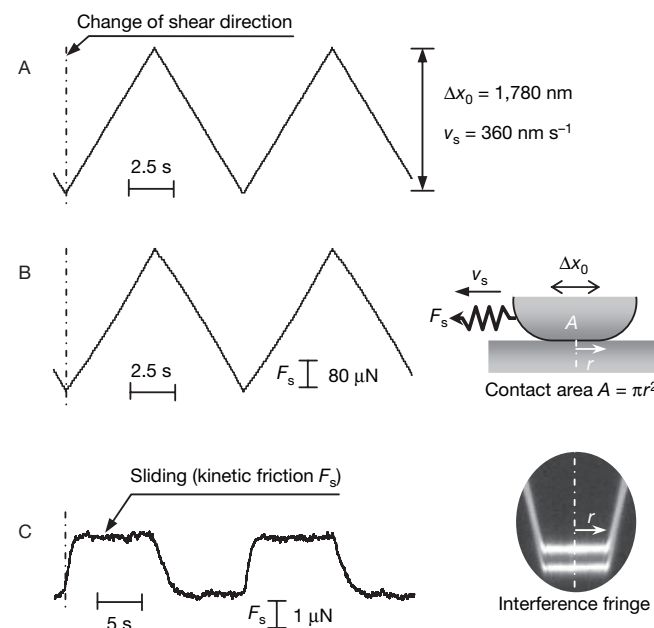


Figure 2 | Characteristics of measured friction between two DDunAB-coated surfaces in adhesive contact. Trace A, back-and-forth shear amplitude Δx_0 applied with time to the top surface; traces B and C, the corresponding shear force F_s transmitted across the surfactant layers. Trace B, dry air, showing rigid coupling of the surfaces, that is, the frictional force exceeds the shear force up to the maximal applied shear amplitude. Trace C is a typical transmitted shear force versus time trace between the two surfactant layers (10 min) in adhesive contact after addition of pure water. The lower right-hand insets illustrate the surface force balance configuration in shear¹¹ and the region over which friction was measured from the corresponding fringe image taken from an actual experiment.

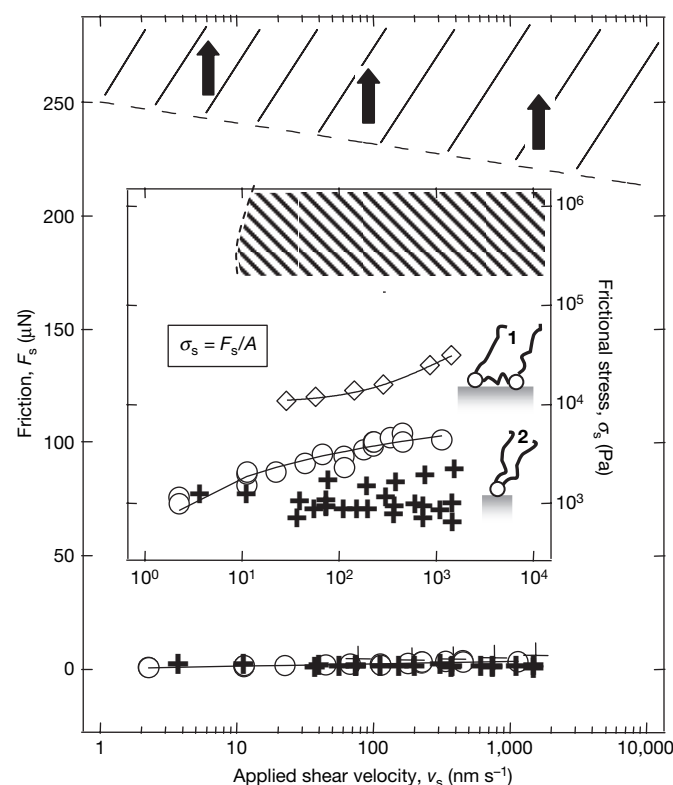


Figure 3 | Dependence of friction force F_s between surfactant layers in initial contact in water on the sliding velocity v_s . F_s is measured from traces such as B and C in Fig. 2. The friction forces in dry air (before adding water) at all velocities and for all surfactants are greater than the upper hatched region. Open circles indicate DDunAB layers (10 min), measured following approach to adhesion across water. Crosses indicate DDunAB layers (10 min) that were in adhesive contact before immersion in water and throughout the friction measurements. The inset shows the corresponding frictional sliding stress $\sigma_s = F_s/A$, where the contact area A between the two surfaces is measured from the flat portion of the optical fringes (as inset in Fig. 2). The data symbols in the inset for DDunAB (schematic 2) correspond to those in the main figure. (The data are identical in the inset and the main figure: the logarithmic scale of the inset highlights the variation of the data relative to that of the linear scale of the main figure.) Open diamonds in the inset indicate frictional sliding stress for the Gemini surfactant (see text; schematic 1). The dry air contact separation between the Gemini-monolayer-bearing surfaces is 1.3 ± 0.3 nm. The hatched area shows, for comparison, the range of frictional sliding stresses from several other comparable experiments^{2,3,10,16} in air (dry, humid or with organic vapour). The solid curves are a guide to the eye.

frictional sliding stress $\sigma_s \approx 2 \times 10^7 \text{ N m}^{-2}$, which is within the range of the frictional stress regime in dry air, but is too high, by some orders of magnitude, to account for the low friction under water, for which $\sigma_s = (1\text{--}5) \times 10^3 \text{ N m}^{-2}$ (inset to Fig. 3). The clear implication is that while, in air, both adhesion and sliding occur at the same surfactant–surfactant interface, the situation under water is different: The weakest adhesion (and therefore separation of the surfaces on pull-off) occurs at the surfactant–surfactant interface, but the sliding under water must occur at a much more lubricated interface: the one between the surfactant headgroups and the substrate. This unexpected result is, however, consistent with estimates of the magnitude of the adhesion at the two interfaces (Methods).

Second, to eliminate the possibility that the sliding may nonetheless have been occurring at the mid-plane, lubricated by hydrated polar headgroups of surfactant molecules that had undergone an overturning or ‘flip-flop’ transition^{23,24} under water, we repeated the frictional experiments using a different procedure. Following measurements in dry air, the surfaces were brought into adhesive contact at $D = D_0 = 2.7 \pm 0.4 \text{ nm}$ as before. On subsequent addition of water, the contact separation increased by $0.6 \pm 0.4 \text{ nm}$, very similar to the swelling $\Delta D_0 = 0.51 \pm 0.4 \text{ nm}$ noted earlier for monolayers which were immersed in water before contact. Frictional forces were then measured as above, taking care not to separate the contacting regions or expose them directly to water at any time during the shear. The frictional stresses measured in this way are shown as the crosses (data points) in Fig. 3. They reveal a comparable (or even slightly

lower) frictional stress than when water is introduced before contact. With this procedure the possibility of any overturning of surfactants^{23,24} and consequent hydration of the surfactant–surfactant interface is suppressed, so we conclude that the slip-plane must indeed be at the surfactant–substrate interface.

Finally, we measured the friction using a boundary lubricant of chemical structure homologous to DDunAB above, but with different architecture, so that it contacts the surface not only via its polar headgroups but also via hydrophobic moieties. We used the symmetric Gemini surfactant $\text{CH}_3(\text{CH}_2)_{11}\text{N}^+(\text{CH}_3)_2\text{Br}^-(\text{CH}_2)_6\text{N}^+(\text{CH}_3)_2\text{Br}^-(\text{CH}_2)_{11}\text{CH}_3$, which contacts the mica via its two polar ($\text{N}^+(\text{CH}_3)_2$) headgroups, as well as via the intervening $(\text{CH}_2)_6$ spacer, as revealed by neutron reflectometry and indicated by schematic 1 in the inset to Fig. 3 (R.K.T. and P. X. Li, unpublished data; see also Supplementary Information). Typical sliding frictional stresses using these surfactants immersed in water are summarized in the inset to Fig. 3. They are lower than for boundary lubrication in air, but are higher than for the DDunAB molecules. This is in line with our expectation that only the polar headgroups of the surfactant layers can become hydrated and provide lubrication. The DDunAB layers have a relatively more hydrated interface with the substrate (schematic 2 in the inset to Fig. 3), and therefore slide more easily than the other surfactant, for which only part of the surfactant/substrate interface is hydrated, the rest being covered by the hydrophobic C_6 alkyl groups—which cannot be hydrated—in close proximity to the surface.

In summary, we have shown that the frictional stress between sliding surfaces coated with amphiphilic surfactant layers (boundary lubricants) immersed in water may be reduced by 1–2 orders of magnitude or more, relative to its value in dry air. We attribute this to the shift of the slip plane during frictional sliding, from the surfactant–surfactant midplane interface (in air) to the much better lubricated surfactant–substrate interfaces (under water). Lubrication at the latter interfaces is mediated, we believe, by the fluid hydration sheaths surrounding the surfactant polar headgroups at the substrate, in analogy to the lubricating effect provided by hydrated ions between compressed sliding surfaces²¹. This unexpected scenario of crossover of the slip plane is supported by several observations: pre-adhering the surfaces before their immersion in water results in similarly low friction, indicating that fluidization of the surfactant–surfactant interface cannot be responsible for the lubrication. In addition, the large adhesion hysteresis both in air and in water indicates adhesion occurs at a different interface (between the surfactants) to that of sliding (at the substrate–surfactant interface). Finally, using chemically homologous surfactants of different architecture can result in much higher friction, as expected when the surfactant–substrate interface is only partly hydrated.

Our results may have implications for the lubrication of biomedical devices and microelectromechanical systems currently limited by friction and wear: for example, in the design of boundary lubricant molecules that will optimize the extent of interfacial hydration. The results suggest applications in living systems that are in aqueous environments: their lubrication may also be mediated by hydration shells (in contrast to conjectures that a classical boundary lubrication mechanism is active⁸), and may point to more efficient treatments for osteoarthritis²⁵.

METHODS

Boundary layer preparation. After calibrating in air contact, the mica surfaces glued onto cylindrical silica lenses (Epon 1004, Shell) were dismounted and immersed for different periods from 10–120 min in the surfactant solution, 0.3–3 mM concentration ($\sim 10\text{--}100$ -fold critical micelle concentration) at $\sim 25^\circ \text{C}$ in water (purified via the RiOs5-MilliQ Gradient A10 system; resistivity is $>18.2 \Omega \text{ cm}$ and total organic content $\leq 4 \text{ p.p.b.}$). In the figure legends, layers prepared at these different incubation times are identified as ‘10 min’ and so on. After withdrawal from the surfactant solution, the surfaces were rinsed thoroughly with water, and dried in a desiccator in the presence of P_2O_5 for $\sim 15 \text{ h}$ before being remounted in the surface force balance (in the presence of P_2O_5).

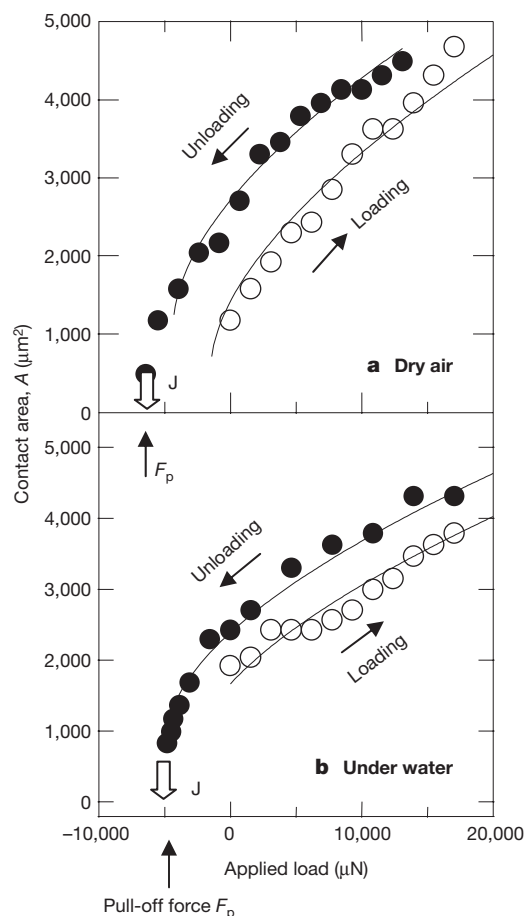


Figure 4 | Loading–unloading profiles of contacting DDunAB (10 min) layers showing variation of contact area A with load L . **a**, In dry air; **b**, under water. The solid lines are fits to the Johnson–Kendall–Roberts expression (see Methods) using best-fit values of the adhesion energy γ on loading (γ_L) and unloading (γ_U), and yield differences $\Delta\gamma = (\gamma_U - \gamma_L)$ in the range $20 \pm 2 \text{ mJ m}^{-2}$ for both cases (differences in absolute pull-off forces are due to different unperturbed radii of curvature R).

The double surfactant layer thickness is 2.2 ± 0.3 nm relative to air, to which 5 ± 3 Å was added to account for the thin adsorbed water/gas layer known to desorb in water²¹, which presumably comes off on immersion in the surfactant solution.

Adhesion hysteresis. The contact area A between the surfaces is varied (and monitored) by changing the load L during compression or decompression runs to yield the data in Fig. 4, from which the adhesion energy γ may be extracted by fitting the $A(L)$ variation to the Johnson-Kendall-Roberts contact mechanics expression¹⁵:

$$A = \pi \{ (R/K) [L + 6\pi R\gamma + (12\pi R\gamma L + (6\pi R\gamma)^2)^{1/2}] \}^{2/3} \quad (0)$$

Here $K = 9.5 \times 10^9$ N m⁻² is the effective modulus of the mica/glue combination. The difference between the value of γ on loading (γ_L) and its value on unloading (γ_U) gives the adhesion energy hysteresis $\Delta\gamma$.

Adhesive forces at the different interfaces. We assume the monolayers detach intact on pull-off so that surfactant molecules remain within their layers. The hydrophobic adhesive force $f_{\text{surf-surf}}$ per surfactant molecule at the surfactant-surfactant interface may be written as¹⁴ $f_{\text{surf-surf}} = [\partial E / \partial D]_{D=0}$ where $E \approx (\gamma s^2) e^{-D/h}$; here $\gamma \approx 35$ mJ m⁻² is the surface energy at the interface between the hydrophobic tail and the water (from the pull-off forces in Fig. 1), $s^2 \approx 50$ Å² is the mean area per surfactant molecule, and $h \approx 1$ nm is the decay length for the hydrophobic attraction¹⁴. This yields $f_{\text{surf-surf}} = (\gamma s^2)/h \approx 2 \times 10^{-11}$ N. The adhesive force $f_{\text{surf-substr}}$ per surfactant molecule attached by a positively charged polar head at the negatively charged substrate may be estimated as $f_{\text{surf-substr}} = -e^2/(4\pi\epsilon_0\epsilon_r x^2)$, where $x \approx 4$ Å is the charge separation and e is the electronic charge. The effective dielectric constant for layers of water comparable in thickness l to a hydration layer is much less than that of the bulk, and has been both calculated²⁶ and evaluated from experiment²⁷ as $\epsilon \approx 5-10$ for $l < 1$ nm. This gives $f_{\text{surf-substr}} \approx (1-2) \times 10^{-10}$ N, an order of magnitude or so larger than $f_{\text{surf-surf}}$ (this estimate assumes the hydration layers detach with the polar head-groups, to be replaced by water at the mica surface). This is consistent with pull-off on separation of the surfaces occurring at the surfactant-surfactant interface, even though sliding occurs at the substrate surfaces.

Received 10 May; accepted 24 August 2006.

1. Bowden, F. P. & Tabor, D. D. *The Friction and Lubrication of Solids* Vols I and II (Oxford Univ. Press, Oxford, 1950, 1964).
2. Briscoe, B. J. & Evans, D. C. B. The shear properties of Langmuir-Blodgett layers. *Proc. R. Soc. Lond. A* **380**, 389–407 (1982).
3. Yoshizawa, H., Chen, Y.-L. & Israelachvili, J. Fundamental mechanisms of interfacial friction. 1. Relation between adhesion and friction. *J. Phys. Chem.* **97**, 4128–4140 (1993).
4. Zhang, L., Li, L., Chen, S. & Jiang, S. Measurements of friction and adhesion for alkyl monolayers on Si(111) by scanning force microscopy. *Langmuir* **18**, 5448–5456 (2002).
5. Grant, L. M. & Tiberg, F. Normal and lateral forces between lipid covered solids in solutions: correlation between layer packing and structure. *Biophys. J.* **82**, 1373–1385 (2002).
6. Suda, H. Origin of friction derived from rupture dynamics. *Langmuir* **17**, 6045–6047 (2001).
7. Zilberman, S., Persson, B. N. J. & Nitzan, A. Theory and simulations of squeeze-out dynamics in boundary lubrication. *J. Chem. Phys.* **115**, 11268–11277 (2001).
8. Hills, B. A. Boundary lubrication *in vivo*. *Proc. Inst. Mech. Eng. H* **214**, 83–94 (2000).

9. Sarma, A. V., Powell, G. L. & LaBerge, M. Phospholipid composition of articular cartilage boundary lubricant. *J. Orthop. Res.* **19**, 671–676 (2001).
10. Drummond, C., Israelachvili, J. & Richetti, P. Friction between two weakly adhering boundary lubricated surfaces in water. *Phys. Rev. E* **67**, 066110 (2003).
11. Klein, J. & Kumacheva, E. Simple liquids confined to molecularly thin layers. I. Confinement-induced liquid to solid phase transitions. *J. Chem. Phys.* **108**, 6996–7009 (1998).
12. McGillivray, D. J., Thomas, R. K., Rennie, A. R., Penfold, J. & Sivia, D. S. Ordered structures observed at the silicon-water interface for di-chain cationic surfactants. *Langmuir* **19**, 7719–7726 (2003).
13. Derjaguin, B. V., Churaev, N. V. & Muller, V. M. *Surface Forces* Ch. 4 and 8 (Consultants Bureau, New York, 1987).
14. Christenson, H. K. & Claesson, P. M. Direct measurements of the forces between hydrophobic surfaces in water. *Adv. Colloid Interf. Sci.* **91**, 391–436 (2001).
15. Johnson, K. L., Kendall, K. & Roberts, A. D. Surface Mechanics and the contact of elastic solids. *Proc. R. Soc. Lond. A* **324**, 301–313 (1971).
16. Yamada, S. & Israelachvili, J. N. Friction and adhesion hysteresis of fluorocarbon surfactant monolayer-coated surfaces measured with the surface forces apparatus. *J. Phys. Chem. B* **102**, 234–244 (1998).
17. Chen, Y. L. & Israelachvili, J. N. Effects of ambient conditions on adsorbed surfactant and polymer monolayers. *J. Phys. Chem.* **96**, 7752–7760 (1992).
18. Williams-Darby, S., Thomas, R. K., Castro, M. A. & Becerro, A. Structure of complexes of Vermiculite intercalated by cationic surfactants. *J. Colloid Interf. Sci.* **256**, 314–324 (2002).
19. Wennerstrom, H., Persson, N. O. & Lindman, B. in *Colloidal Dispersions and Micellar Behaviour* (ed. Mittal, K. L.) Ch. 18 253–269 (ACS Symp. Ser. 9, American Chemical Society, Washington, 1975).
20. Hamnerius, Y., Lundström, I., Paulsson, L. E., Fontell, K. & Wennerström, H. Dielectric properties of lamellar lipid water phases. *Chem. Phys. Lipids* **22**, 135–140 (1978).
21. Raviv, U. & Klein, J. Fluidity of bound hydration layers. *Science* **297**, 1540–1543 (2002).
22. Raviv, U., Giasson, S., Gohy, J.-F., Jerome, R. & Klein, J. Lubrication by charged polymers. *Nature* **425**, 163–165 (2003).
23. Chen, Y. L. E., Gee, M. L., Helm, C. A., Israelachvili, J. N. & McGuigan, P. M. Effects of humidity on the structure and adhesion of amphiphilic monolayers on mica. *J. Phys. Chem.* **93**, 7057–7059 (1989).
24. Langmuir, I. Overturning and anchoring of monolayers. *Science* **87**, 493–500 (1938).
25. Vecchio, P., Thomas, K. & Hills, B. A. Surfactant treatment for osteoarthritis. *Rheumatology* **38**, 1020–1021 (1999).
26. Stern, H. A. & Feller, S. E. Calculation of the dielectric permittivity profile for a nonuniform system. *J. Chem. Phys.* **118**, 3401–3412 (2003).
27. Teschke, O., Ceotto, G. & de Souza, E. F. Dielectric exchange force: a convenient technique for measuring the interfacial water relative permittivity profile. *Phys. Chem. Chem. Phys.* **3**, 3761–3768 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Chen and P.X. Li for experimental assistance, and I. Dunlop, J. Israelachvili, S. Perkin and B. Roux for comments. Financial support from the EPSRC (UK) and from the Israel Science Foundation (J.K.) is acknowledged with thanks.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.K. (jacob.klein@chem.ox.ac.uk or jacob.klein@weizmann.ac.il).

One-to-one coupling of glacial climate variability in Greenland and Antarctica

EPICA Community Members*

Precise knowledge of the phase relationship between climate changes in the two hemispheres is a key for understanding the Earth's climate dynamics. For the last glacial period, ice core studies^{1,2} have revealed strong coupling of the largest millennial-scale warm events in Antarctica with the longest Dansgaard–Oeschger events in Greenland^{3–5} through the Atlantic meridional overturning circulation^{6–8}. It has been unclear, however, whether the shorter Dansgaard–Oeschger events have counterparts in the shorter and less prominent Antarctic temperature variations, and whether these events are linked by the same mechanism. Here we present a glacial climate record derived from an ice core from Dronning Maud Land, Antarctica, which represents South Atlantic climate at a resolution comparable with the Greenland ice core records. After methane synchronization with an ice core from North Greenland⁹, the oxygen isotope record from the Dronning Maud Land ice core shows a one-to-one coupling between all Antarctic warm events and Greenland Dansgaard–Oeschger events by the bipolar seesaw⁶. The amplitude of the Antarctic warm events is found to be linearly dependent on the duration of the concurrent stadial in the North, suggesting that they all result from a similar reduction in the meridional overturning circulation.

The glacial climate in the North Atlantic region is characterized by rapid shifts from cold stadial to warmer interstadial conditions^{3,4,9}. Greenland temperatures during these Dansgaard–Oeschger (D–O) events rise by 8–16 °C (refs 10, 11) within a few decades followed by a less rapid temperature decline back to stadial conditions. In contrast, glacial climate in the circum-Antarctic region exhibits slower millennial changes with smaller temperature amplitudes of only 1–3 °C (refs 1, 12, 13). After synchronization of Greenland and Antarctic ice core records^{1,2} using the global atmospheric change in CH₄ concentrations, a conspicuous phase relationship between the largest Antarctic warmings (A1–A7; ref. 1) and the longest D–O events was observed with the south warming during the stadial conditions in the north, and starting to cool as soon as the D–O warming set in. This bipolar seesaw pattern was explained by changes in the heat and freshwater flux connected to the Atlantic Meridional Overturning Circulation (MOC), where a stronger MOC leads to increased drainage of heat from the Southern Ocean heat reservoir^{6,7}.

In principle, an interhemispheric climate coupling by the bipolar seesaw should also apply for all the short D–O events. However, to what extent this concept is also able to explain the higher-frequency climate variability in Antarctic ice cores remained unclear (as discussed for example, in ref. 14 and references therein). Here we report

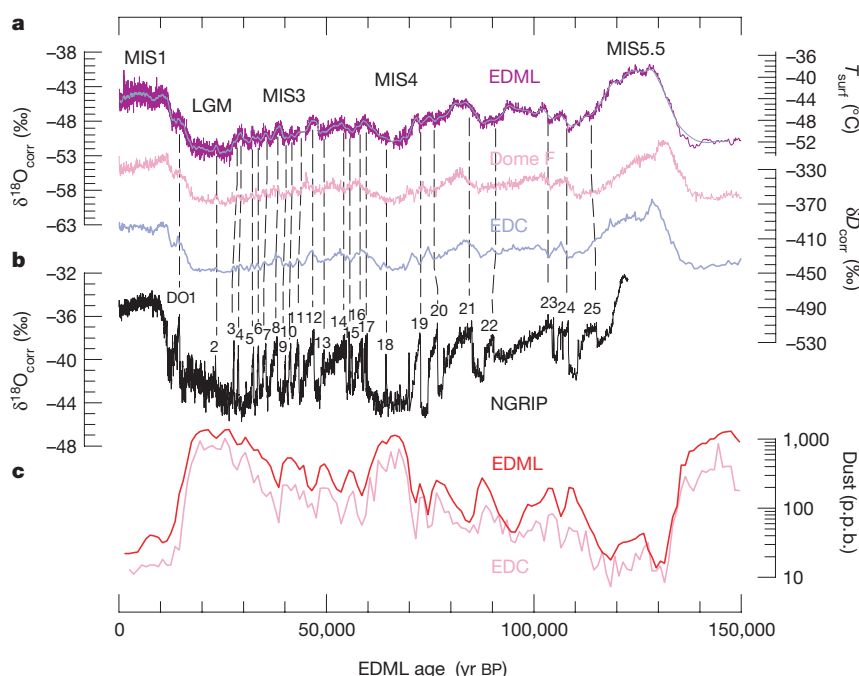


Figure 1 | Antarctic stable isotope records show synchronous millennial variations during the last glacial, whereas rapid variations are encountered in Greenland. a, EDML $\delta^{18}\text{O}$ record (purple, 0.5-m resolution; grey, 15-m running mean) after sea level and upstream correction (see Supplementary Information) over the past 150 kyr. The record shows features similar to those of the EDC¹² (blue) and the Dome F¹³ (pink) isotope records but with more fine structure during MIS3 and MIS4. We note that EDML and EDC are plotted on the new common EDC3 timescale (see Supplementary Information) while Dome F is plotted on its individual timescale. The temperature axis on the right side indicates approximate surface temperatures at EDML as derived from the spatial $\delta^{18}\text{O}$ /temperature gradient (see Supplementary Information). **b**, $\delta^{18}\text{O}$ record of the NGRIP ice core (grey)⁹. **c**, Mineral dust records of the EDML (red) and EDC¹² (pink) ice cores at 1,000-yr resolution; these dust records were used for synchronization of the cores.

*A full list of authors and their affiliations appears at the end of the paper.

on the climate record over the last glacial cycle from a new ice core drilled within the European Project for Ice Coring in Antarctica (EPICA) in the interior of Dronning Maud Land, hence denoted EDML, at 75° S, 0° E, 2,892 m.a.s.l. (metres above sea level), with a recent accumulation rate of 6.4 cm water equivalent (w.e.) per year¹⁵. This site was chosen to complement the long EPICA Dome C (EDC, 75° S, 123° E, 3,233 m.a.s.l., 2.5 cm w.e. yr⁻¹) record¹², because EDML is the first deep ice core in the Atlantic sector of the Southern Ocean region¹⁶ and thus located near the southern end of the bipolar seesaw. The snow accumulation at EDML is two to three times higher than at other deep drilling sites on the East Antarctic plateau, so higher-resolution atmosphere and climate records can be obtained for the last glacial period, making the EDML core especially suitable for studying decadal-to-millennial climate variations in Antarctica.

In Fig. 1 the EDML $\delta^{18}\text{O}$ record as proxy for local temperature on the ice sheet is shown in 0.5-m resolution (equivalent to 15–30 yr during the marine isotope stage MIS3 and 100–150 yr during MIS5) after correction for upstream and glacial–interglacial ice sheet altitude effects (see Supplementary Information). The overall pattern closely resembles that recorded in most Antarctic ice cores previously covering this time period^{12,13,17}. Also, very similar dust profiles (Fig. 1) are encountered at EDML and EDC, related to parallel changes in climate conditions in the Patagonian dust source region common to both cores¹⁸. Despite the high correlation of the EDML $\delta^{18}\text{O}$ and the EDC δD record over the last 150,000 yr ($r^2 = 0.94$ for 250-yr averages) some distinct differences exist. In the penultimate warm period (MIS5.5) the EDML $\delta^{18}\text{O}$ record indicates temperatures about 4–5 °C higher than those of the Holocene, in line with other ice cores from the East Antarctic plateau^{12,13,17}. However, $\delta^{18}\text{O}$ at EDML exhibits persistently higher $\delta^{18}\text{O}$ values over the entire duration of MIS5.5 while other ice cores on the East Antarctic plateau show a substantial drop after an initial climate optimum^{12,13}. We note that this difference is not due to the altitude corrections applied to the EDML $\delta^{18}\text{O}$ record (see Supplementary Information), because a similar temporal evolution during MIS5.5 is also seen in the uncorrected data. Instead, a smaller cooling at EDML in the course of MIS5.5 compared to EDC and Dome Fuji is consistent with marine sediment records from the Atlantic sector of the Southern Ocean revealing persistently warmer summer sea surface temperatures

and a reduced winter sea ice cover throughout MIS5.5 (ref. 19). This suggests that there were regional differences in temperature and sea ice evolution during this period for the Atlantic and Indian Ocean sector.

The most outstanding feature of the high-resolution EDML record is the pronounced millennial variability during the glacial. As indicated by the dashed lines in Fig. 1 each of the warming episodes in Antarctica can be related to a corresponding D–O event, but only synchronization of the age scales allows us to assign them unambiguously and to pinpoint the phase relationship between climate changes in Greenland and Antarctica. To do this, the EDML core has been synchronized (see Supplementary Information) to the layer counted NGRIP ice core^{20,21} over MIS3, using high-resolution CH_4 profiles over the last 55 kyr from the NGRIP, GRIP and GISP2 ice cores^{1,11}. The synchronized $\delta^{18}\text{O}$ records are shown in Fig. 2. Also plotted is the CH_4 synchronized $\delta^{18}\text{O}$ record from the Byrd ice core¹ and new high-resolution δD data from EDC²² which closely resemble the temperature variability found at EDML during MIS3 and support an Antarctic-wide interpretation of these fluctuations. The higher glacial snow accumulation at EDML (~ 3 cm w.e. yr⁻¹) compared to that at EDC, Dome Fuji or Vostok (~ 1.4 cm w.e. yr⁻¹) implies a CH_4 synchronization two to three times better than at those sites. The synchronization uncertainty for MIS3 ranges from 400 to 800 yr for all events in the EDML record, making the synchronization error for EDML always much smaller than the duration of the events themselves.

This is important, because this allows an unequivocal one-to-one assignment not only of the well-known large warm events in Antarctica (A1, A2 and so on) but of each single isotope maximum indicated in Fig. 2 with a corresponding D–O event in the north. Although the exact timing of the temperature maxima relative to the stadial/interstadial transitions cannot be discerned more precisely than the synchronization error, it is evident that each Antarctic warming starts significantly before the respective D–O event. In addition, a synchronization of the stable water isotope records of the GRIP and EDC ice cores using the ^{10}Be production anomaly around 41,000 yr BP, which constrains the in-phase relationship of the onset of D–O 10 and the respective Antarctic δD maximum to better than 200 yr (ref. 23), supports our CH_4 match. Accordingly, we suggest a new Antarctic Isotope Maximum (AIM) nomenclature in Fig. 2

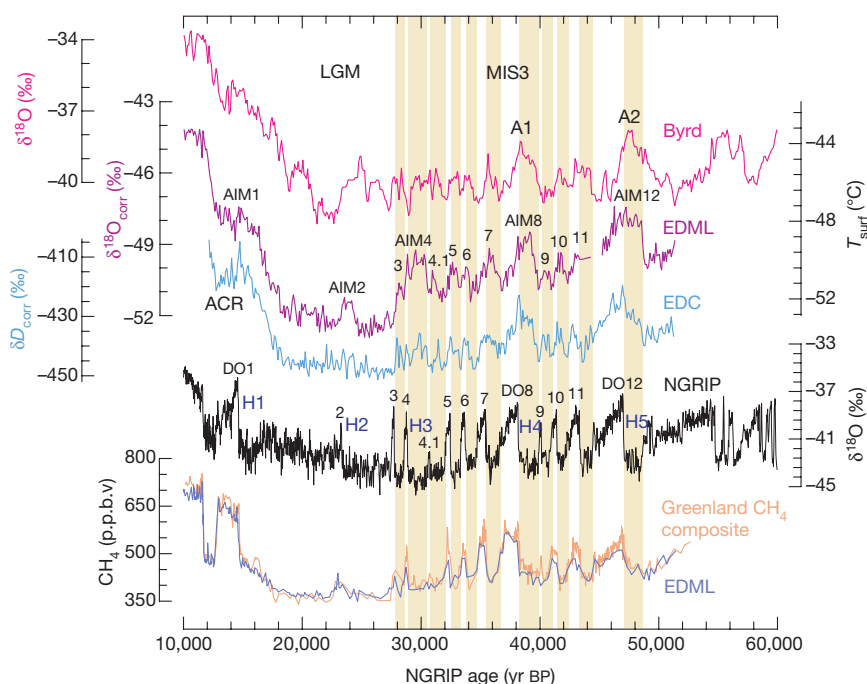


Figure 2 | Methane synchronization of the EDML and the NGRIP records reveals a one-to-one assignment of each Antarctic warming with a corresponding stadial in Greenland. Displayed are 100-yr averages during MIS3 in the EDML, EDC²⁶ and Byrd¹ ice core for the time interval 10–60 kyr BP in comparison with the NGRIP $\delta^{18}\text{O}$ record from Northern Greenland⁹. All records are CH_4 synchronized and given on the new GICC05 age scale for the NGRIP ice core, which has been derived by counting annual layers down to 41 kyr and by a flow model for older ages^{9,21}. Yellow bars indicate the Greenland stadial periods that we relate to respective Antarctic temperature increases. The approximate timing of Heinrich layers in North Atlantic sediments is indicated as well²⁷. The y axis on the right side indicates approximate temperature changes at EDML based on the modern spatial gradient between $\delta^{18}\text{O}$ and temperature.

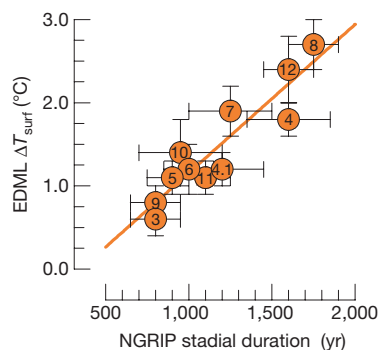


Figure 3 | Amplitudes of Antarctic warmings show a linear relationship ($r^2 = 0.85$) with the duration of the accompanying stadial in Greenland during MIS3. The amplitude was determined from the Antarctic $\delta^{18}\text{O}$ maximum to the preceding minimum of each event; the stadial duration is defined by the interval between the midpoint of the stepwise temperature change at the start and end of a stadial on the extended GICC05 age scale^{9,21}. Error bars reflect the estimated uncertainty in the definition of the maxima and minima in $\delta^{18}\text{O}$ at EDML and in the duration of the concurrent stadial period. Numbers indicate the corresponding AIMs and D-O events.

which reflects the connection of southern warming to reduced oceanic heat transport into the North Atlantic during stadials. The timing and duration of the AIMs relative to D-O events is also indirectly supported by the comparison of changes in deep-water masses linked to Antarctic Bottom Water formation and Atlantic surface water changes, as archived in sediment records offshore of Portugal²⁴.

Most striking is the varying amplitude of the AIMs, which is linearly dependent on the duration of stadials in the north, as shown in Fig. 3. The only significant deviation from this linear relationship during MIS3 is AIM4, in which the error in the stadial duration estimate is quite large. We conclude that the duration of a reduced MOC—and, hence, the duration of the warming period in the Southern Ocean—determines the amount of heat accumulated in the Southern Ocean heat reservoir, strongly supporting the general applicability of the thermal bipolar seesaw⁶ concept within the range of stadial events encountered during MIS3. We note that for longer cessations of the MOC a new equilibrium temperature in the Southern Ocean would be reached and the warming would eventually have to cease. This linear relationship also implies that the Antarctic warming rate—and thus the heat flux from the Southern Ocean to the Atlantic—is similar for all warming events during MIS3. If we assume the same spatial configuration of the overturning cell for cold intervals in MIS3, this would suggest that the strength of the MOC is approximately constant for all stadials, challenging the notion of different overturning rates²⁵ for stadials in which massive iceberg discharges into the North Atlantic (the so-called Heinrich events in Fig. 2: H1–H5) occurred compared to stadials without Heinrich events. Note however, that the stadials before D-O 8 and D-O 12 in which Heinrich events occurred were the longest and the related Antarctic warmings the largest. This may be due to the longer time needed to mix away the large freshwater anomalies during Heinrich events. There is, however, a less clear relationship for other Heinrich events. Comparison of the millennial climate variability during MIS3 at EDML and EDC shows no significant difference in the amplitude of the isotopic change in the Atlantic and Indian Ocean sectors of the Southern Ocean. This implies a uniform ocean heat reservoir controlling temperature changes at both sites and reflects the rapid mixing of the Southern Ocean by the Antarctic Circumpolar Current.

In the EDML $\delta^{18}\text{O}$ record a major warm event (AIM2, connected to D-O 2) is seen during the Last Glacial Maximum, which cannot be clearly identified in the EDC core but is present in the Dome F record (Fig. 1). AIM2 also shows a decrease in high-resolution mineral dust concentrations at EDC, as do all the other AIMs²⁶. We therefore

conclude that AIM2 is a warm event comparable to the other AIMs in MIS3 but is not sufficiently resolved in the EDC record owing to its lower accumulation. The corresponding D-O 2 event in the North Atlantic is preceded by the longest cold period in the NGRIP record (Fig. 2) and accordingly, a higher temperature amplitude of AIM2 is to be expected if the same bipolar seesaw concept holds as for D-O events during MIS3. However, sea level and temperature conditions were significantly different during the Last Glacial Maximum, potentially affecting the spatial configuration and strength of the overturning cell in the North Atlantic. The fact that AIM2 is only 2,000 yr long suggests that the strength of the MOC was not significantly reduced for the entire cold period in the North, but collapsed only about 1,000 yr before D-O 2, which would be in line with significant iceberg discharge depositing ice-rafted debris in the North Atlantic during H2 (ref. 27).

In summary, a strong interhemispheric coupling of all bipolar climate variations during MIS3 via the MOC is supported by the new high-resolution $\delta^{18}\text{O}$ record from EDML indicating that Antarctic warming rates and potentially also overturning rates have been similar for all events in MIS3. The question of what triggers the switch from stadial to interstadial conditions remains. Transitions in the strength of the MOC and its effect on the Atlantic Southern Ocean heat exchange are simulated in response to changes in the North Atlantic freshwater balance^{7,8}; however, the origin of such variations in freshwater input are still not ascertained for all individual D-O events. In addition, large iceberg discharge from the Laurentide ice sheet does not systematically coincide with either the onset or the end of stadials^{27,28}. Recently, the potential role of a change in Southern Ocean sea-ice cover for reinstalling a stronger MOC has been identified for the onset of the Bølling/Allerød warming^{29,30}. The intrinsic feedback of a reduced sea-ice cover in the Southern Ocean during AIMs, followed by a delayed onset of deep-water formation in the North, could potentially explain the interhemispheric climate coupling seen in our records during MIS3.

Received 26 May; accepted 22 September 2006.

- Blunier, T. & Brook, E. J. Timing of millennial-scale climate change in Antarctica and Greenland during the last glacial period. *Science* **291**, 109–112 (2001).
- Blunier, T. *et al.* Asynchrony of Antarctic and Greenland climate change during the last glacial period. *Nature* **394**, 739–743 (1998).
- Johnsen, S. J. *et al.* Irregular glacial interstadials recorded in a new Greenland ice core. *Nature* **359**, 311–313 (1992).
- Bond, G. *et al.* Correlations between records from North Atlantic sediments and Greenland ice. *Nature* **365**, 143–147 (1993).
- McManus, J. F., Oppo, D. W. & Cullen, J. L. A 0.5-million-year record of millennial climate variability in the North Atlantic. *Science* **283**, 971–975 (1999).
- Stocker, T. F. & Johnsen, S. J. A minimum thermodynamic model of the bipolar seesaw. *Paleoceanography* **18**, art. no. 1087 (2003).
- Knutti, R., Flückiger, J., Stocker, T. F. & Timmermann, A. Strong hemispheric coupling of glacial climate through freshwater discharge and ocean circulation. *Nature* **430**, 851–856 (2004).
- Ganopolski, A. & Rahmstorf, S. Rapid changes of glacial climate simulated in a coupled climate model. *Nature* **409**, 153–158 (2001).
- North Greenland Ice Core Project members. High resolution climate record of the northern hemisphere reaching into the last interglacial period. *Nature* **431**, 147–151 (2004).
- Landais, A. *et al.* Quantification of rapid temperature change during DO event 12 and phasing with methane inferred from air isotopic measurements. *Earth Planet. Sci. Lett.* **225**, 221–232 (2004).
- Huber, C. *et al.* Isotope calibrated Greenland temperature record over Marine Isotope Stage 3 and its relation to CH_4 . *Earth Planet. Sci. Lett.* **245**, 504–519 (2006).
- EPICA community members. Eight glacial cycles from an Antarctic ice core. *Nature* **429**, 623–628 (2004).
- Watanabe, O. *et al.* Homogeneous climate variability across East Antarctica over the past three glacial cycles. *Nature* **422**, 509–512 (2003).
- Roe, G. H. & Steig, E. J. Characterization of millennial-scale climate variability. *J. Clim.* **17**, 1929–1944 (2004).
- Oerter, H. *et al.* Accumulation rates in Dronning Maud Land, Antarctica, as revealed by dielectric-profiling measurements of shallow firn cores. *Ann. Glaciol.* **30**, 27–34 (2000).
- Reijmer, C. H., van den Broeke, M. R. & Scheele, M. P. Air parcel trajectories to five deep drilling locations on Antarctica, based on the ERA-15 data set. *J. Clim.* **15**, 1957–1968 (2002).

17. Petit, J. R. *et al.* Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429–436 (1999).
18. Basile, I. *et al.* Patagonian origin of glacial dust deposited in East Antarctica (Vostok and Dome C) during glacial stages 2, 4 and 6. *Earth Planet. Sci. Lett.* **146**, 573–589 (1997).
19. Bianchi, C. & Gersonde, R. The Southern Ocean surface between Marine Isotope Stage 6 and 5d: Shape and timing of climate changes. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **187**, 151–177 (2002).
20. Rasmussen, S. O. *et al.* A new Greenland ice core chronology for the last glacial termination. *J. Geophys. Res.* **111**, D06102 (2006).
21. Andersen, K. K. *et al.* The Greenland ice core chronology 2005, 15–42 kyr. Part 1: Constructing the time scale. *Quat. Sci. Rev.* (in the press).
22. Stenni, B. *et al.* A late-glacial high-resolution site and source temperature record derived from the EPICA Dome C isotope records (East Antarctica). *Earth Planet. Sci. Lett.* **217**, 183–195 (2003).
23. Raisbeck, G., Yiou, F. & Jouzel, J. Cosmogenic ^{10}Be as a high resolution correlation tool for climate records. *Geochim. Cosmochim. Acta* **66**, abstr. A623 (2002).
24. Shackleton, N. J., Hall, M. A. & Vincent, E. Phase relationships between millennial-scale events 64,000–24,000 years ago. *Paleoceanography* **15**, 565–569 (2000).
25. Rahmstorf, S. Ocean circulation and climate during the past 120,000 years. *Nature* **419**, 207–214 (2002).
26. Röthlisberger, R. *et al.* Dust and sea-salt variability in central East Antarctica (Dome C) over the last 45 kyr and its implications for southern high-latitude climate. *Geophys. Res. Lett.* **29**, article no. 1963 (2002).
27. Bond, G. & Lotti, R. Iceberg discharges into the North Atlantic on millennial time scales during the last glaciation. *Science* **267**, 1005–1010 (1995).
28. de Abreu, L., Shackleton, N. J., Joachim Schönfeld, J., Hall, M. & Chapman, M. Millennial-scale oceanic climate variability off the Western Iberian margin during the last two glacial periods. *Mar. Geol.* **196**, 1–20 (2003).
29. Knorr, G. & Lohmann, G. Southern Ocean origin for the resumption of Atlantic thermohaline circulation during deglaciation. *Nature* **424**, 532–536 (2003).
30. Stocker, T. F. & Wright, D. G. Rapid transitions of the ocean's deep circulation induced by changes in surface water fluxes. *Nature* **351**, 729–732 (1991).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work is a contribution to the European Project for Ice Coring in Antarctica (EPICA), a joint European Science Foundation/European Commission scientific programme, funded by the EU (EPICA-MIS) and by national contributions from Belgium, Denmark, France, Germany, Italy, the Netherlands, Norway, Sweden, Switzerland and the UK. The main logistic support was provided by IPEV and PNRA (at Dome C) and AWI (at Dronning Maud Land).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H. F. (hufischer@awi-bremerhaven.de).

EPICA Community Members (listed in alphabetical order): C. Barbante^{1,2}, J.-M. Barnola³, S. Becagli⁴, J. Beer⁵, M. Bigler^{6,7}, C. Boutron³, T. Blunier⁶, E. Castellano⁴, O. Cattani⁸, J. Chappellaz³, D. Dahl-Jensen⁷, M. Debret³, B. Delmonte⁹, D. Dick¹⁰, S. Falourd⁸, S. Faria^{10,11}, U. Federer⁶, H. Fischer¹⁰, J. Freitag¹⁰, A. Frenzel¹⁰, D. Fritzsche¹², F. Fundel¹⁰, P. Gabrielli^{2,3}, V. Gaspari¹, R. Gersonde¹⁰, W. Graf¹³, D. Grigoriev¹⁴, I. Hamann¹⁰, M. Hansson¹⁵, G. Hoffmann⁸, M. A. Hutterli^{6,16}, P. Huybrechts^{10,17}, E. Isaksson¹⁸, S. Johnsen⁷, J. Jouzel⁸, M. Kaczmarzka¹⁸, T. Karlin¹⁵, P. Kaufmann⁶, S. Kipfstuhl¹⁰, M. Kohn¹⁰, F. Lambert⁶, Anja Lambrecht¹⁰, Astrid Lambrecht¹⁰, A. Landais⁸, G. Lawer¹⁰, M. Leuenberger⁶, G. Littot¹⁶, L. Loulergue³, D. Lüthi⁶, V. Maggi⁹, F. Marino⁹, V. Masson-Delmotte⁸, H. Meyer¹², H. Miller¹⁰, R. Mulvaney¹⁶, B. Narcisi¹⁹, J. Oerlemans²⁰, H. Oerter¹⁰, F. Parrenin³, J.-R. Petit³, G. Raisbeck²¹, D. Raynaud³, R. Röthlisberger¹⁶, U. Ruth¹⁰, O. Rybak¹⁰, M. Severi⁴, J. Schmitt¹⁰, J. Schwander⁶, U. Siegenthaler⁶, M.-L. Siggaard-Andersen⁷, R. Spahni⁶, J. P. Steffensen⁷, B. Stenni²², T. F. Stocker⁶, J.-L. Tison²³, R. Traversi⁴, F. Valero-Delgado¹⁰, M. R. van den Broeke²⁰, R. S. W. van de Wal²⁰, D. Wagenbach²⁴, A. Wegner¹⁰, K. Weiler⁶, F. Wilhelms¹⁰, J.-G. Winther¹⁸ & E. Wolff¹⁶

¹Department of Environmental Sciences, University Ca' Foscari of Venice, ²Institute for the Dynamics of Environmental Processes-CNR, Dorsoduro 2137, 30123 Venice, Italy. ³Laboratoire de Glaciologie et Géophysique de l'Environnement (LGGE), CNRS-UJF, BP96 38402 Saint-Martin-d'Hères cedex, France. ⁴Department of Chemistry, University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Florence, Italy. ⁵EAWAG, PO Box 611, 8600 Dübendorf, Switzerland. ⁶Climate and Environmental Physics, Physics Institute, University of Bern, Sidlerstrasse 5, 3012 Bern, Switzerland. ⁷Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, 2100 Copenhagen OE, Denmark. ⁸Laboratoire des Sciences du Climat et de l'Environnement (LSCE/IPSL), CEA-CNRS-UVSQ, CE Saclay 91191, Gif sur Yvette, France. ⁹Environmental Sciences Department, University of Milano Bicocca, Piazza della Scienza 1, 20126 Milano, Italy. ¹⁰Alfred-Wegener-Institute for Polar and Marine Research, Columbusstrasse, D-27568 Bremerhaven, Germany. ¹¹Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, 04103 Leipzig, Germany. ¹²Alfred Wegener Institute for Polar and Marine Research, Research Unit Potsdam, Telegrafenberg A 43, 14473 Potsdam, Germany. ¹³GSF National Center for Environment and Health, Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany. ¹⁴University College London, Gower Street, London WC1E 6BT, UK. ¹⁵Department of Physical Geography and Quaternary Geology, Stockholm University, 106 91 Stockholm, Sweden. ¹⁶British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 0ET, UK. ¹⁷Departement Geografie, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium. ¹⁸Norwegian Polar Institute, 9296 Tromsø, Norway. ¹⁹ENEA, C. R. Casaccia, Via Anguillarese 301, 00060 Roma, Italy. ²⁰Utrecht University, Institute for Marine and Atmospheric Research, PO Box 80005, 3508 TA Utrecht, The Netherlands. ²¹CSNSM/IN2P3/CNRS, Bat. 108, 91405 Orsay, France. ²²Department of Geological, Environmental and Marine Sciences, University of Trieste, Via E. Weiss 2, 34127 Trieste, Italy. ²³Département des Sciences de la Terre, Université Libre de Bruxelles, CP160/03, 1050 Brussels, Belgium. ²⁴Institute for Environmental Physics, University of Heidelberg, INF229, 69120 Heidelberg, Germany.

An exceptional Devonian fish from Australia sheds light on tetrapod origins

John A. Long^{1,2,3}, Gavin C. Young², Tim Holland^{1,3}, Tim J. Senden⁴ & Erich M. G. Fitzgerald^{1,3}

The transition from fishes to tetrapods was one of the most dramatic events in the evolution of vertebrates, but many pivotal fossils are incomplete, resulting in gaps in the data that are used for phylogenetic reconstruction. Here we present new observations from the most complete, acid-prepared Devonian tetrapodomorph fish yet discovered, *Gogonasus*^{1,2}, which was previously placed just crownward of *Kenichthys* and rhizodontids^{3,4}, the most primitive taxa on the tetrapod lineage. Unexpectedly, *Gogonasus* shows a mosaic of plesiomorphic and derived tetrapod-like features. Whereas the braincase and dermal cranial skeleton exhibit generalized morphologies with respect to *Eusthenopteron*⁵ or *Panderichthys*⁶, taxa that are traditionally considered to be phylogenetically close to tetrapods^{7,8}, the presence of a deeply invaginated, wide spiracle, advanced internal spiracular architecture and near-horizontal hyomandibula are specialized features that are absent from *Eusthenopteron*⁹. Furthermore, the pectoral fin skeleton of *Gogonasus* shares several features with that of *Tiktaalik*, the most tetrapod-like fish¹⁰. A new phylogenetic analysis places *Gogonasus* crownward of *Eusthenopteron* as the sister taxon to the Elpistegalia. Aspects of the basic tetrapod limb skeleton and middle ear architecture can now be traced further back within the tetrapodomorph radiation.

Recently, the discovery of transitional fossil forms of advanced tetrapod-like fishes^{6,10–12} and primitive tetrapods^{13–15} from the Devonian period have provided important insights into the evolution of the first land vertebrates¹⁶. For almost a century, the anatomy of *Eusthenopteron*, which is considered to be a fairly derived tetrapodomorph fish^{5,7–9}, has served as a basis for comparison with early tetrapods.

Many complete specimens of *Eusthenopteron* have been studied, and it has been universally positioned by many workers^{3–9,11} as an intermediate form between the generalized cosmine-covered tetrapodomorph fishes (such as *Osteolepis*) and the more advanced tetrapod-like elpistogalians^{6,10,11}. *Kenichthys*⁴, the most generalized sarcopterygian fish directly on the tetrapodomorph lineage, is known only from isolated skulls and cranial dermal bones, so we lack knowledge of its paired fins, endoskeletal girdles, gill arches, and overall body form to contribute to the understanding of early character development in the lineage. New anatomical data on basal tetrapodomorph fishes is therefore vital to clarify the incipient stages that led to the origin of the tetrapod body plan.

The new specimen described here (NMV P221807, Museum Victoria, Melbourne) was discovered in 2005 in a limestone concretion of the Lower Frasnian Gogo Formation of Western Australia (dated as ~384–380 million years old). The Gogo fish fauna (more than 44 taxa) is widely acknowledged^{17–19} for its perfect three-dimensional preservation, and *Gogonasus* is its only known tetrapodomorph. Previous specimens^{1,2} lacked the detailed preservation of

the spiracular region, complete visceral skeleton, body and fins seen in this exceptional new example, study of which has been enhanced by high resolution X-ray tomography after acid preparation. The conspicuously large spiracular opening (Fig. 1a–c) is proportionally similar to those recently reconstructed for *Panderichthys*⁹ and *Tiktaalik*¹¹. The pectoral fin endoskeleton of *Gogonasus* is described here for the first time (Fig. 2), the new specimen being the only known Devonian fish that shows a complete acid-prepared pectoral limb. There are some surprising similarities to the recently described pectoral fin in the advanced elpistostegalian *Tiktaalik*¹⁰. As such features could indicate homoplasy between *Gogonasus* and early tetrapods, we present a revised character analysis to determine whether the new anatomical information supports a more crownward position for *Gogonasus* in the stem-tetrapod phylogeny.

The articulated cranium of *Gogonasus* (Fig. 1a, b) shows the natural gaps between the boundary of the skull-roof and the cheek unit. The most conspicuous feature is the ventrally down-turned cosmine-covered lamina of the tabular bone, which forms a posterior externally open wall to the spiracular opening (Fig. 1d, e). The spiracle is a very small slit-like opening in *Eusthenopteron*⁵, progressively larger in elpistostegalians, and forms a wide otic notch in early tetrapods⁹. No previously described tetrapodomorph fish shows such a large spiracular opening, or a downward facing dermal lamina forming a posterior wall to the spiracular chamber, so the condition in *Gogonasus* is highly unusual. This indicates that spiracular breathing might have evolved independently in some stem tetrapodomorphs. The shape of the spiracular chamber has been restored from the perfectly preserved palatoquadrate with the distinct ridge on the entopterygoid delineating the ventral boundary of the spiracular region (the posterior wall of the eustachian tube) as in *Eusthenopteron* (see Supplementary Information for details). *Panderichthys* lacks such a ridge on the entopterygoid, but still has a horizontally oriented spiracular chamber⁹. The shallower angle of the spiracular chamber margin in *Gogonasus* (Fig. 1g) is a perfect intermediate morphology between the deeper spiracular chamber of *Eusthenopteron* (Fig. 1h) and the almost horizontal chamber of *Panderichthys*⁹ (Fig. 1f). However, in having the entopterygoid located lateral to the ventral opening of the spiracular tract, the condition in *Panderichthys* is more derived than either *Eusthenopteron* or *Gogonasus*.

Both hyomandibulae are well preserved, with similar morphology to the previously described *Gogonasus* specimens². Compared to the relatively long and slender hyomandibula of *Eusthenopteron*, that of *Gogonasus* has similar relative proportions (proximal:distal section ratio) but with a slightly thicker, more robust proximal shaft. The more horizontal angle of the *Gogonasus* hyomandibula compared to that of *Eusthenopteron* is confirmed by the new specimen, except that the opercular process does not meet the skull roof bones, instead attaching at its ventral end to an elevated, dorsolaterally facing

¹Museum Victoria, P O Box 666, Melbourne, Australia 3001. ²Department of Earth and Marine Sciences, The Australian National University, Canberra, Australia 0200. ³School of Geosciences, Monash University, Clayton, Victoria, Australia 3800. ⁴Department of Applied Mathematics, Research School of Physical Sciences and Engineering, The Australian National University, Canberra, Australia 0200.

opercular bone. The dorsal position of the hyomandibula conforms more closely to that in elpistostegalians than to those in tristichopterids and more generalized tetrapodomorphs.

The pectoral fin skeleton of *Gogonasus* (Fig. 2) shows not only the three-dimensional form of the endoskeletal elements, but also preserves spacing between elements as filled by cartilage in life. The humerus (Fig. 2a–d, f, g) has a relatively flat cross-section and broad, strap-like caput humeri. An unusual short ventral process between

the metapodial facets and the entepicondyle is also present in *Tiktaalik*¹⁰. The entepicondyle (Fig. 2a, c, d, f) is significantly smaller, relative to the length of the humerus shaft, than those of rhizodontids and tristichopterids (Fig. 3c), and the two subrectangular facets of the ulna and radius are offset at an angle similar to that of *Tiktaalik*, in having a forward-reaching radial joint (Fig. 2f). The radius (Fig. 2f–i) is an elongate, tapering bone, unlike the significantly broader or distally flared hour-glass elements seen in tristichopterids and

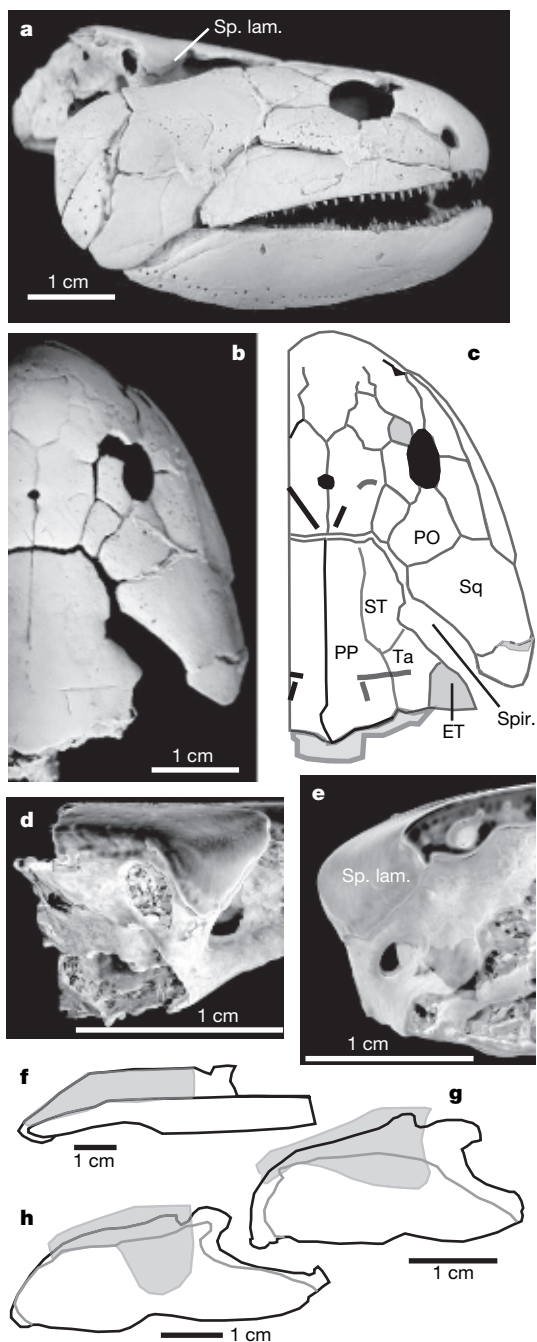


Figure 1 | Cranial features of *Gogonasus andrewsae*, NMV P221807. **a**, Skull in lateral view. **b**, **c**, skull in dorsal view. **d**, **e**, X-ray micro-tomograms showing the spiracular region in lateral (**d**) and anterior (**e**) views. **f**–**h**, Restorations of palatoquadrates showing the spiracular chamber (shaded) for *Panderichthys* (**f**), *Gogonasus* (**g**) and *Eusthenopteron* (**h**); panels **f** and **h** are based on ref. 9. ET, extratemporal bone; PO, postorbital; PP, postparietal; Spir, spiracular opening; Sp. lam., down-turned lamina of tabular for spiracular chamber; Sq, squamosal; ST, supratemporal; Ta, tabular. All bones whitened with ammonium chloride.

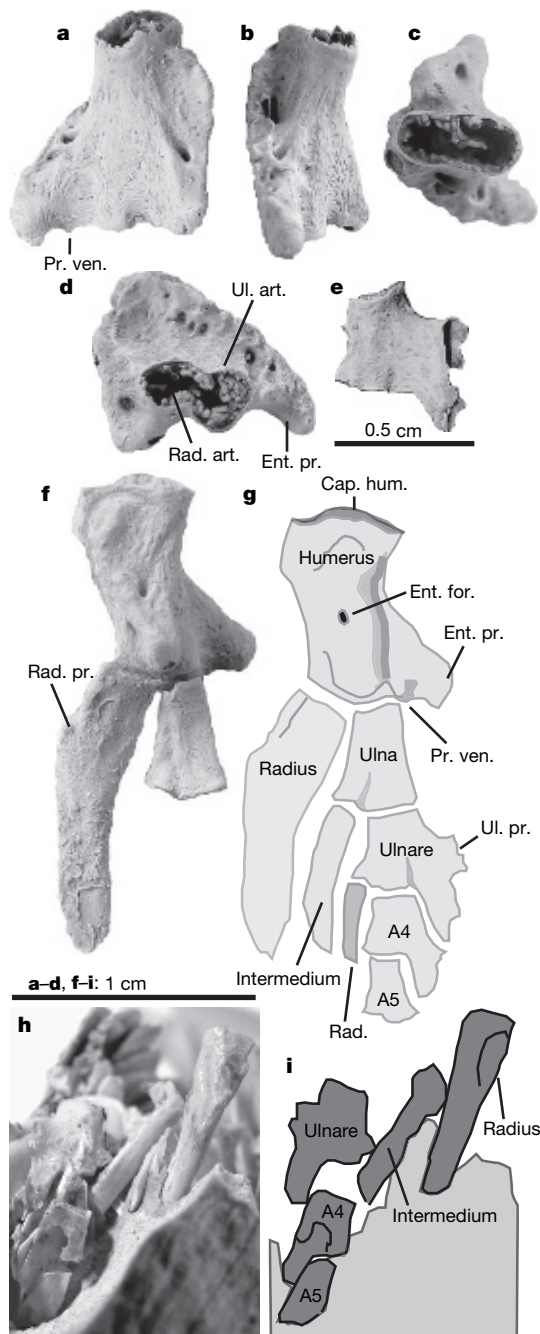


Figure 2 | Pectoral fin skeleton of *Gogonasus andrewsae*, NMV P221807. **a**–**d**, Right humerus in dorsal (**a**), mesial (**b**), proximal (**c**) and distal (**d**) views. **e**, Right ulnare in ventral view. **f**, Left humerus, ulna and radius articulated in dorsal view. **g**, Reconstructed left pectoral fin in dorsolateral view. **h**, Left pectoral distal elements. **i**, Key to **h**. A4, A5, distal mesomeres of the pectoral fin; Cap. hum., caput humeri; Ent. for., entepicondylar foramen; Ent. pr., entepicondyle; Pr. ven., ventral process on entepicondyle; Rad., presumed radial element; Rad. art., radial articulation; Rad. pr., radial process; Ul. pr., postaxial process of ulnare. All (except **g**) whitened with ammonium chloride. Scale bar, 1 cm for **a**–**d**, **f**–**i**; 0.5 cm for **e**.

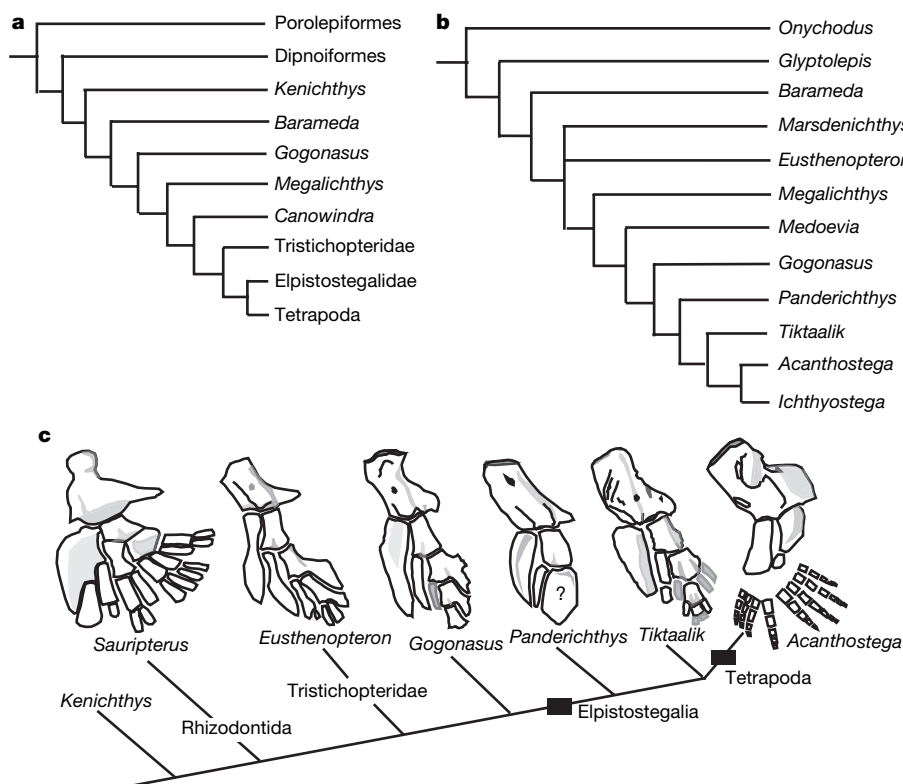


Figure 3 | Phylogenetic position of *Gogonasus*. **a**, Previous phylogenetic position of *Gogonasus* (based on refs 4, 7). **b**, New position of *Gogonasus* based on our PAUP analysis of 103 characters (see Supplementary Information for protocols). **c**, Position of *Gogonasus* relative to other

tetrapodomorph fishes, showing pectoral fin endoskeleton comparisons (diagrams after ref. 10). We note that new information on the correct pectoral fin arrangement of *Panderichthys* is forthcoming (P. E. Ahlberg, personal communication).

rhizodontids (Fig. 3c). It also has a well defined radial process (Fig. 2f). The elongate ulna (Fig. 2f, g) has two distal articulations, one large one for the ulnare and a small one for a thin, tapering intermedium. The ulnare (Fig. 2g–i) shows a well developed postaxial process, as in *Eusthenopteron*, except that in *Gogonasus* the process does not extend very far distally as a flared lamina. The ulnare articulates distally with a broad, flat mesomere (A4) and a smaller anterior facet indicates that smaller radial elements were also present. The A4 element has a single distal, broad A5 element articulating with it (Fig. 2g–i), in contrast to the slender paired elements seen in *Eusthenopteron* and the rhizodontids. This condition approaches that of *Tiktaalik*, which has one large element and three narrow elements in this position. Such features can be interpreted as either generalized (plesiomorphic) for *Gogonasus* and elpistostegalians, or shared apomorphies that unite them, and as such would exclude the rhizodontids and tristichopterids from the higher clade.

Our new cladistic analysis (Fig. 3b) places *Gogonasus* crownwards of tristichopterids as a sister taxon to elpistostegalians, based mainly on its spiracular morphology and pectoral fin skeleton (Fig. 3c). Supporting this is the fact that elpistostegalian fishes still retain primitive rhombic scales, as seen in many basal sarcopterygian fishes. The rounded scales of tristichopterids and rhizodontids²⁰ characteristically possess a median boss on their inner surface, consistent with the suggestion that these groups were divergent from the lineage containing *Gogonasus* and elpistostegalians, although the current analysis lacks evidence to unite them in a separate clade.

Gogonasus is a marine tetrapodomorph fish with cosmine-covered rhombic scales and dermal bones, in this respect resembling *Kenichthys* from China, the earliest member of the tetrapodomorph clade in which an incipient internal nostril or choana has been identified⁴. Thus, this key tetrapod innovation evolved as early as the Emsian (late Early Devonian), and the new data from *Gogonasus* indicate that the basic tetrapod pectoral fin pattern, and the broad

spiracular opening as a precursor to tetrapod middle ear architecture, might also have originated further back than previously thought within the tetrapodomorph radiation.

Our new phylogeny replaces the tristichopterid *Eusthenopteron* as the typical fish model for the fish–tetrapod transition. It also raises the question of what environment the immediate stem group of the elpistostegalians inhabited. The marine environment inhabited by *Gogonasus* is in accord with the marginal marine environments of some elpistostegalians (*Panderichthys*, *Elpistostegia*, *Tiktaalik*) and the tetrapod *Tulerpeton*. Such observations support a model in which the first tetrapods, like their immediate piscine sister taxa, were capable of marine dispersal, thus explaining the widespread global distribution achieved shortly after their first appearance in the late Frasnian.

Finally, we note that strata of similar age to those producing *Gogonasus*, *Panderichthys* and *Tiktaalik*, or slightly younger, have yielded tetrapod jaws in Australia and China^{21,22}, and also trackways attributed to two different unknown tetrapods on the Gondwana supercontinent²³. This indicates that the initial radiation of tetrapods from elpistostegalian fishes, with evidence currently confined to the northern hemisphere landmass of Euramerica, was probably an extremely rapid global event. Migration of some Middle–Late Devonian fishes from a Gondwana place of origin to Euramerica has been well documented²⁴. With *Gogonasus* now positioned phylogenetically close to elpistostegalians, we suggest that the current lack of fossil evidence for elpistostegalian fishes in Gondwana could be due to poor sampling. We are now exploring areas in Australia with undescribed Devonian sarcopterygian fish remains to test this hypothesis.

METHODS

Preparation. The acid immersion procedure (in 10% acetic acid for 2 days, washed in running water for a day, air-dried, all exposed bones then hardened

with Mowital B30 consolidant dissolved in pure ethanol) was repeated over 4 months to fully extract the anterior two-thirds of the specimen, with the tail section embedded in an acrylic resin slab to retain articulation.

Phylogenetic analysis. The matrix includes 103 morphological characters in 12 taxa, of which 9 were based on fairly complete specimens with few missing codings. All characters were treated as unordered, and were weighted equally. Character data entry and formatting was performed in MacClade (version 4.05). Our new analysis is based on the data matrix for the recently described, and now best known, elpistogalian, *Tiktaalik*^{10,11}, with some new characters added and some recoded (see Supplementary Information for details and the data matrix). We use the recently described basal sarcopterygian *Onychodus* from Gogo¹⁹ together with *Glyptolepis* for our outgroups. For rhizodontids, new material of the Australian form *Barameda*²⁰ provided more complete data on the primitive fin skeleton (poorly known in *Gooloogongia*, as used in previous analyses⁴). The character analysis was performed using PAUP 4.1. The Supplementary Information provides more details on the phylogenetic analysis and the X-ray micro-tomography used to study the specimen.

Received 4 June; accepted 11 September 2006.

Published online 18 October 2006.

- Long, J. A. A new osteolepidid fish from the Upper Devonian Gogo Formation of Western Australia. *Recs. W. A. Mus.* **12**, 361–377 (1985).
- Long, J. A. *et al.* Osteology and functional morphology of the osteolepidiform fish *Gogonassus* Long, 1985, from the Upper Devonian Gogo Formation, Western Australia. *Recs. W. A. Mus. Suppl.* **57**, 1–89 (1997).
- Chang, M. M. & Zhu, M. A new Middle Devonian osteolepidid from Qujing, Yunan. *Mem. Australas. Ass. Palaeontol.* **15**, 183–198 (1993).
- Zhu, M. & Ahlberg, P. E. The origin of the internal nostril of tetrapods. *Nature* **432**, 94–97 (2004).
- Jarvik, E. *Basic Structure and Evolution of Vertebrates* (Academic, London, 1980).
- Vorobjeva, E. & Schultze, H.-P. in *Origins of the Higher Groups of Tetrapods: Controversy and Consensus* (eds Schultze, H.-P. & Trueb, L.) 68–109 (Cornell Univ. Press, Ithaca, New York, 1991).
- Ahlberg, P. E. & Johanson, Z. Osteolepiformes and the ancestry of tetrapods. *Nature* **395**, 792–794 (1998).
- Chang, M. M. & Yu, X. Re-examination of the relationships of Middle Devonian osteolepids—fossil characters and their interpretations. *Am. Mus. Novitates* **3189**, 1–20 (1997).
- Brazeau, M. & Ahlberg, P. E. Tetrapod-like middle ear architecture in a Devonian fish. *Nature* **439**, 318–321 (2006).
- Shubin, N. *et al.* The pectoral fin of *Tiktaalik rosae* and the origin of the tetrapod limb. *Nature* **440**, 764–771 (2006).
- Daeschler, E. *et al.* A Devonian tetrapod-like fish and the evolution of the tetrapod body plan. *Nature* **440**, 757–763 (2006).
- Johanson, Z. & Ahlberg, P. E. Devonian rhizodontids and tristichopterids (Sarcopterygii) from East Gondwana. *Trans. R. Soc. Edinb. Earth Sci.* **92**, 43–74 (2001).
- Ahlberg, P. E. *et al.* A near tetrapod from the Baltic Middle Devonian. *Palaeontology* **43**, 533–548 (2000).
- Ahlberg, P. E. *Elginerpeton pancheni* and the earliest tetrapod clade. *Nature* **373**, 420–425 (1995).
- Ahlberg, P. E. *et al.* The first tetrapod finds from the Devonian (Upper Famennian) of Latvia. *Phil. Trans. R. Soc. Lond. B* **343**, 303–328 (1994).
- Clack, J. *Gaining Ground. The Origin and Evolution of Tetrapods* 1–369 (Indiana Univ. Press, Bloomington, 2002).
- Ahlberg, P. E. Fossil fishes from Gogo. *Nature* **337**, 511–512 (1989).
- Janvier, P. *Early Vertebrates* 1–393 (Clarendon, Oxford, 1996).
- Andrews, S. M. *et al.* The structure of the sarcopterygian *Onychodus jandemarrai* n. sp. from Gogo, Western Australia: with a functional interpretation of the skeleton. *Trans. R. Soc. Edinb. Earth Sci.* **96**, 197–307 (2006).
- Holland, T. *et al.* A new species of *Barameda* (Rhizodontida) and heterochrony in the rhizodontid pectoral fin. *J. Vert. Paleontol.* (in the press).
- Campbell, K. S. W. & Bell, M. W. A primitive amphibian from the Late Devonian of New South Wales. *Alcheringa* **1**, 369–381 (1977).
- Zhu, M. *et al.* First Devonian tetrapod from Asia. *Nature* **420**, 760–761 (2002).
- Warren, J. W. & Wakefield, N. A. Trackways of tetrapod vertebrates from the Upper Devonian of Victoria, Australia. *Nature* **228**, 469–470 (1972).
- Young, G. C. An articulated phyllolepid fish (Placodermi) from the Devonian of central Australia: implications for non-marine connections with the Old Red Sandstone continent. *Geol. Mag.* **142**, 173–186 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. E. Ahlberg for discussion of the new material and access to new tomographic data on the hyomandibula of *Eusthenopteron*, and for unpublished information on the pectoral fin of *Panderichthys*, and M. I. Coates for discussion of the material. J.A.L. and G.C.Y. are supported by the Australian Research Council, T.J.S. acknowledges support under an ARC Research Fellowship, and the X-ray Tomography Facility is supported by an ARC LIEF grant. T.H. is supported by a Monash University Dean's Postgraduate Scholarship, and E.M.G.F. is supported by an Australian Postgraduate Award.

Author Contributions The specimen was prepared and described by J.A.L., G.C.Y. and T.H.; T.J.S. conducted X-ray tomography of the specimen including visualization; E.M.G.F. conducted the phylogenetic analyses and study of character evolution with data matrix input from J.A.L., G.C.Y. and T.H.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.A.L. (jl@museum.vic.gov.au).

Retinal repair by transplantation of photoreceptor precursors

R. E. MacLaren^{1,2*}, R. A. Pearson^{3*}, A. MacNeil¹, R. H. Douglas⁴, T. E. Salt⁵, M. Akimoto^{6†}, A. Swaroop^{6,7}, J. C. Sowden³ & R. R. Ali^{1,8}

Photoreceptor loss causes irreversible blindness in many retinal diseases. Repair of such damage by cell transplantation is one of the most feasible types of central nervous system repair; photoreceptor degeneration initially leaves the inner retinal circuitry intact and new photoreceptors need only make single, short synaptic connections to contribute to the retinotopic map. So far, brain- and retina-derived stem cells transplanted into adult retina have shown little evidence of being able to integrate into the outer nuclear layer and differentiate into new photoreceptors^{1–4}. Furthermore, there has been no demonstration that transplanted cells form functional synaptic connections with other neurons in the recipient retina or restore visual function. This might be because the mature mammalian retina lacks the ability to accept and incorporate stem cells or to promote photoreceptor differentiation. We hypothesized that committed progenitor or precursor cells at later ontogenetic stages might have a higher probability of success upon transplantation. Here we show that donor cells can integrate into the adult or degenerating retina if they are taken from the developing retina at a time coincident with the peak of rod genesis⁵. These transplanted cells integrate, differentiate into rod photoreceptors, form synaptic connections and improve visual function. Furthermore, we use genetically tagged post-mitotic rod precursors expressing the transcription factor Nrl (ref. 6) (neural retina leucine zipper) to show that successfully integrated rod photoreceptors are derived only from immature post-mitotic rod precursors and not from proliferating progenitor or stem cells. These findings define the ontogenetic stage of donor cells for successful rod photoreceptor transplantation.

We assessed the transplantation potential of immature mouse retinal cells, taken from the early postnatal period at the peak of rod photoreceptor genesis (postnatal day (P)1)⁵. At this age, the retinal microenvironment should be favourable to promote the differentiation and integration of transplanted cells within the outer nuclear layer (ONL). Furthermore, transplanted cells have a higher probability of integration if recipient and donor retinas are at equivalent stages of development. Neural retinal cell suspensions ($\sim 2 \times 10^5$ cells per injection) from P1 transgenic mice, expressing a green fluorescent protein (GFP) gene under the control of a chicken β -actin promoter (*Cba-gfp*^{+/+})⁷, were transplanted into the subretinal space of GFP-negative wild-type littermates. Three weeks after transplantation, substantial numbers of cells (10–200 cells per eye; $N = 10$ eyes) had migrated into the recipient neural retina. Most (>95%) were

correctly oriented within the ONL and had morphological features typical of mature photoreceptors (Fig. 1a, b).

As a population of cells in the P1 retina could integrate and differentiate into photoreceptors when transplanted in the immature retina, we transplanted P1 cells ($\sim 8 \times 10^5$ cells per eye) into the subretinal space of adult GFP-negative wild-type mice. By contrast to previous reports^{1,8}, we found that transplanted cells could migrate into the ONL of adult recipients (see also Supplementary Fig. 1). The cells integrated in proportionately similar numbers to those transplanted into the immature retina (300–1,000 cells per eye; $N = 6$) and had the morphological characteristics of mature photoreceptors (Fig. 1b–e). Virtually all integrated cells were rod-like, as expected given the developmental stage of the donor cells^{5,9}. Cone-like profiles were occasionally observed (Fig. 1e). The site of injection seemed to be crucial because, as noted by others², there was no integration within the ONL in either P1 ($N = 12$) or adult ($N = 18$) recipients after intravitreal injections.

Fusion between transplanted and host cells has been proposed as an explanation for the apparent plasticity of stem cells^{10–12}. To evaluate this possibility, we transplanted P1 GFP-positive cells into adult transgenic mice that ubiquitously expressed cyan fluorescent protein (CFP) (*Cba-cfp*^{+/+})¹³. Confocal scans of integrated cells showed that GFP and CFP signals were not co-localized in any of the retinas studied ($N = 8$; Supplementary Fig. 2a). Cell fusion might result in multinuclear cells^{12,14}, but here integrated cells had only a single nucleus. DNA labelling of GFP-positive donor cells with bromodeoxyuridine (BrdU) confirmed that these nuclei originated from donors (Supplementary Fig. 2b). Therefore, it seems unlikely that cell fusion occurred in our experiments.

The population of cells derived from the P1 retina comprises a mixture of proliferating progenitors, post-mitotic precursors and differentiated cells that do not yet express the markers of mature photoreceptors⁵. We sought to determine which of these integrated into the ONL. We examined the developmental time window for obtaining donor cells that successfully integrate after transplantation. Cells were taken from embryonic day (E)11.5, E16.5, P1–P15 or adult GFP-positive donors and transplanted by standardized subretinal injections (see Methods) into adult wild-type recipients. Cells derived from E11.5 retinas, the latest stage that consists almost entirely of proliferating progenitors^{5,9} (Supplementary Fig. 3), survived after transplantation, but in all cases failed to integrate ($N = 12$) (Fig. 2a). Similarly, cells derived from adult retinas survived

¹Division of Molecular Therapy, University College London Institute of Ophthalmology, 11–43 Bath Street, London EC1V 9EL, UK. ²Vitreoretinal Service, Moorfields Eye Hospital, 162 City Road, London EC1V 2PD, UK. ³Developmental Biology Unit, University College London Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK. ⁴Henry Wellcome Laboratory for Vision Sciences, Department of Optometry and Visual Science, City University, Northampton Square, London EC1V 0HB, UK. ⁵Division of Visual Science, University College London Institute of Ophthalmology, 11–43 Bath Street, London EC1V 9EL, UK. ⁶Department of Ophthalmology and Visual Sciences and ⁷Department of Human Genetics, University of Michigan, Ann Arbor, Michigan 48105, USA. ⁸Molecular Immunology Unit, University College London Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK. [†]Present address: Translational Research Center, Kyoto University Hospital, Sakyo-ku, Kyoto 606-8507, Japan.

*These authors contributed equally to this work.

but consistently failed to integrate ($N = 6$). By contrast, cells derived from P1–P7 retinas, which are primarily immature rod precursors, showed robust integration, which was greatest with P3–P5 donors (Fig. 2a). In all cases, a large subretinal collection of viable cells was found at the time of death, indicating that lack of integration was not due to poor cell survival.

The failure of transplanted immature progenitors to integrate was unexpected, but indicates that the properties of these cells change at or after terminal mitosis. To test this, we used BrdU labelling of the adult recipient mice ($N = 12$) to label P1 donor cells that divided after transplantation. Mitotic donor cells survive and continue to

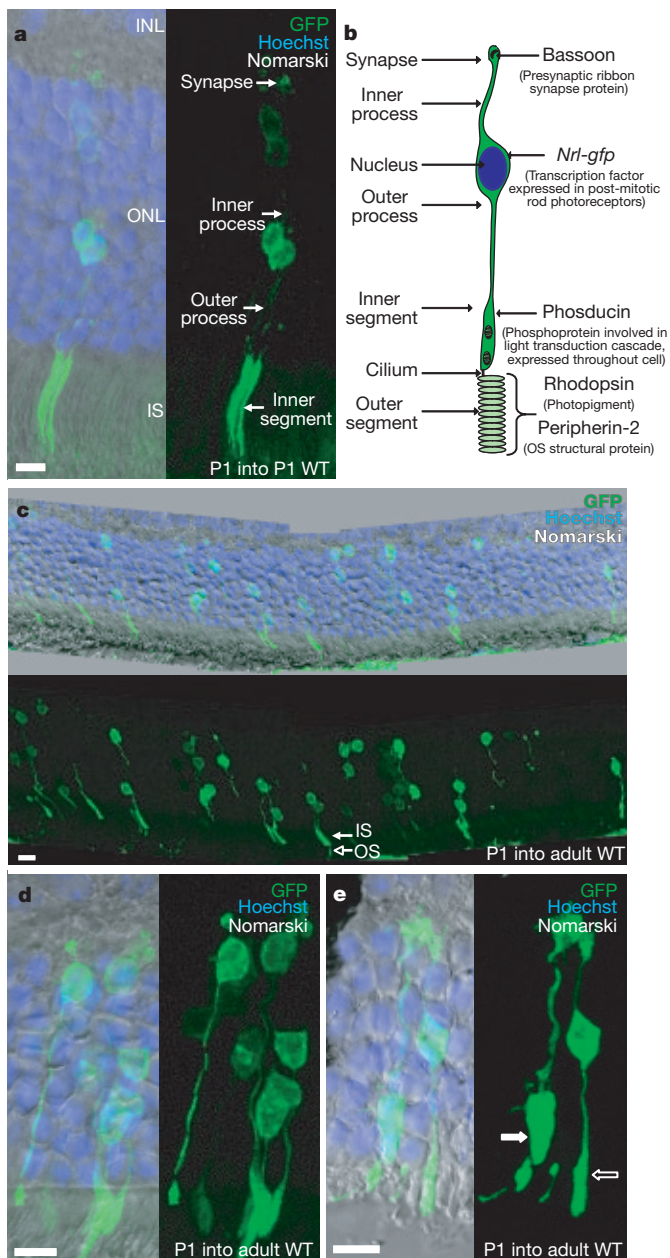


Figure 1 | Integration of P1 retinal cells into immature and adult wild-type recipient retinas. **a**, GFP-positive P1 cells integrated into the ONL of wild-type P1 littermate recipients three weeks after sub-retinal transplantation and developed morphological features typical of mature photoreceptors. **b**, Diagram of mature photoreceptor. **c**, **d**, Low- (**c**) and high- (**d**) magnification images of P1 donor cells integrated into adult wild-type recipients. Note that GFP localization is poor in the outer segments of photoreceptors of these transgenics⁶. **e**, Examples of cells with rod (open arrow) and cone (filled arrow) morphologies. INL, inner nuclear layer; IS, inner segments. Scale bars, 10 μ m.

divide in the subretinal space of the recipient eye (Fig. 2b), but on no occasion did we find BrdU-labelled cells integrated into the recipient retina (Fig. 2c). This implies that the cells that can integrate into the recipient retina are not proliferating progenitors. To confirm the post-mitotic nature of integrating cells, we used a transgenic mouse (*Nrl-gfp*^{+/+})⁶ in which GFP expression is driven by the promoter of *Nrl*, a transcription factor that is specific for post-mitotic rod precursors and that persists in adult rods^{6,15–17} (Supplementary Fig. 1). We used fluorescence-activated cell sorting (FACS) to isolate GFP-positive post-mitotic rod precursors from dissociated P1 *Nrl-gfp*^{+/+} retinas, and transplanted these cells into adult wild-type recipients ($N = 6$). Sorted *Nrl-gfp*^{+/+} cells routinely integrated into the ONL of recipient retinas (Fig. 2d, e) in similar numbers (200–800 cells per eye; $N = 6$) to unsorted transplants (see Methods), indicating that the optimal ontogenetic stage for donor cells for effective rod photoreceptor transplantation corresponds with *Nrl* expression (that is, specification of rod fate).

We then transplanted a population of proliferating progenitor cells from E11.5 *Nrl-gfp*^{+/+} donors, a stage before the onset of *Nrl* expression⁶, and found that these cells failed to integrate into the host retina ($N = 8$). However, they could differentiate to a stage where

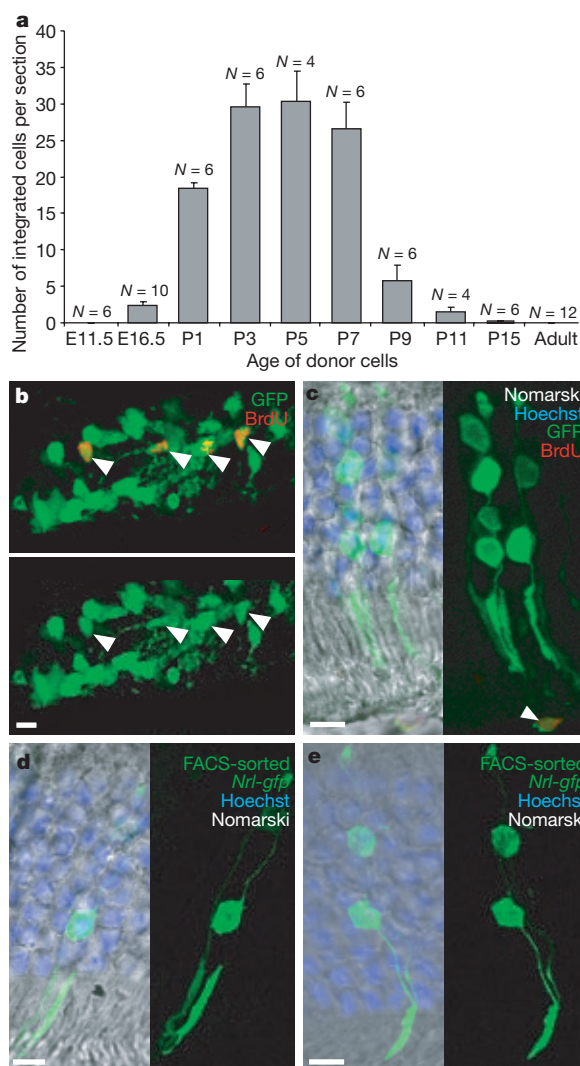


Figure 2 | Optimal ontogenetic stage of donor cells is the post-mitotic photoreceptor precursor. **a**, Number of integrated cells as a function of donor age (mean $\pm$ s.e.m.) after subretinal injection into adult wild-type recipients. **b**, P1 donor cells continued to proliferate in the subretinal space, as indicated by BrdU labelling (red; arrowheads). **c**, Integrated cells were not BrdU labelled. **d**, **e**, FACS-sorted *Nrl-gfp*^{+/+} post-mitotic rod precursor cells readily integrate into the adult ONL. Scale bars, 10 μ m.

both Nrl and rhodopsin (Rho) are expressed, and formed organized rosettes in the subretinal space (Supplementary Fig. 4). The adult retina, therefore, seems to be able to support the survival and differentiation of progenitor cells. However, integration and differentiation of rod photoreceptors is only achieved when cells are transplanted at a later ontogenetic stage.

Integrated cells had the morphological appearance of mature rod photoreceptors. Their identity was confirmed with three additional methods. First, as described above, the restriction of *Nrl-gfp* expression to rods^{6,16,17} provides genetic evidence that most transplanted integrated cells in the ONL are rod photoreceptors. Second, three

weeks after transplantation, integrated cells were immunopositive for phosducin (Fig. 3a) and the photopigment rhodopsin (Fig. 4c), elements of the phototransduction cascade. Integrated cells expressed the ribbon synapse protein bassoon (Fig. 3b) and made synaptic contacts with rod bipolar cells, identified by immunostaining with protein kinase C (Fig. 3c). Thus, integrated cells assemble structural components of the spherule synapse¹⁸, an essential requirement if they are to communicate with the inner retina (Fig. 3c). Finally, we assessed the presence of a subtype of metabotropic glutamate receptor, mGluR8, which is rod-specific and localized to the spherule ribbon synapse^{19,20} (Fig. 3d–f). Application of either glutamate (data not shown) or a specific mGluR8 agonist, (S)-3,4-dicarboxyphenylglycine (DCPG), consistently evoked appropriate decreases in intracellular calcium in both recipient and *Nrl-gfp*^{+/+} integrated cells, and these decreases could be blocked by the specific antagonist (RS)- α -cyclopropyl-4-phosphonophenylglycine (CPPG) (Fig. 3e, f). Conversely, specific agonists of another glutamate receptor, the N-methyl-D-aspartate (NMDA) receptor, which is expressed by other retinal cells but not photoreceptors²⁰, had no effect (data not shown). Together, these findings confirm the identity of transplanted cells that integrate into the ONL as rod photoreceptors that correctly express essential molecules for phototransduction. Furthermore, these cells form synaptic connections with downstream targets and respond to specific, synapse-dependent stimuli in a manner indistinguishable from endogenous photoreceptors.

For cell transplantation to be a viable therapeutic strategy, donor cells must integrate and survive in degenerating retinas and restore visual function. We transplanted P1 *Nrl-gfp*^{+/+} cells into the subretinal space of three mouse models of inherited retinal degeneration; retinal degeneration slow (*rds*), retinal degeneration fast (*rd*) and a rhodopsin knockout (*rho*^{-/-}). Malfunction and degeneration of rods occurs in all these strains and mutations in the corresponding human genes lead to various forms of retinal dystrophy^{21–23}. The *rds* mouse has a null mutation in *peripherin-2*, which is required for the generation of photoreceptor outer segment discs. ONL degeneration begins ~2 weeks after birth, continuing slowly over ~12 months^{24,25}. *Nrl-gfp*^{+/+} donor cells integrated into the retina of adult *rds* mice (*N* = 16) in similar numbers to wild-type recipients (Fig. 4a), remaining viable for at least 10 weeks (the latest time point studied). As predicted, peripherin-2 staining was absent in recipient photoreceptors, but was seen in short outer segments emerging from transplanted cells (Fig. 4a, b). The *rd* mouse retina degenerates rapidly, reducing the ONL to a single layer of predominantly cones by 3 weeks²⁶. Unlike host rods, P1 *Nrl-gfp*^{+/+} cells transplanted into the P1 *rd* mouse retina survived, although with variable morphology owing to the collapse of surrounding tissue (*N* = 8; Supplementary Fig. 5). Degeneration is slower in the *rho*^{-/-} mouse, with the ONL degenerating by ~12 weeks²⁷. P1 *Nrl-gfp*^{+/+} cells transplanted into 4-week-old *rho*^{-/-} mice integrated into the recipient ONL (*N* = 6). Rhodopsin immunostaining was localized to the outer segments (Fig. 4c), as was staining for peripherin-2 after transplantation into the *rds* mouse.

To assess whether transplanted cells were light-responsive and functionally connected to downstream targets, we used two techniques; pupillometry and extracellular field potential recordings from the ganglion cell layer (GCL). We recorded from seven-week-old *rho*^{-/-} mice, which have no functional rod photoreceptors and are insensitive to low light intensities^{28,29}. These mice retain some cone function at early stages and can therefore detect high-intensity stimuli (>0.1 cd s⁻¹ m⁻²)²⁸. *Rho*^{-/-} mice received P1 *Nrl-gfp*^{+/+}; *rho*^{+/+} donor cells in one eye and a sham transplantation of P1 *rho*^{-/-} donor retinal cells in the other, three weeks before assessment.

Light-evoked extracellular field potentials recorded from the GCL of uninjected *rho*^{-/-} mice showed an absence of GCL activity at low light intensities (when rods would be stimulated). Threshold responses (see Methods) were only discernible at a stimulus intensity of 0.052 cd s⁻¹ m⁻² (Fig. 4d); such intensities fall within the range of

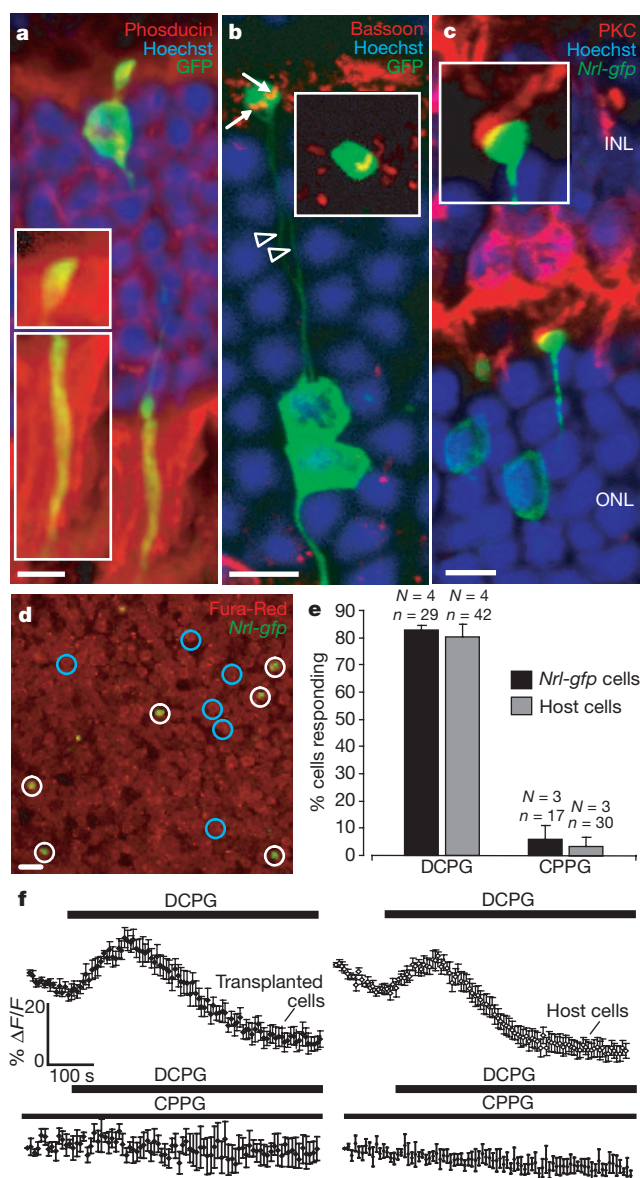


Figure 3 | Photoreceptor identity and synaptic connectivity of integrated cells. **a–c**, Confocal projection images of adult retinal sections stained for phosducin (**a**), bassoon in two adjacent synapses (arrows) from adjacent rods (arrowheads) (**b**), and protein kinase C (PKC) (**c**) (red). Inserts, high-power single confocal sections taken through region of GFP expression only. Scale bars, 10 μ m. **d–f**, Integrated cells respond to stimulation of the rod-specific glutamate receptor, mGluR8, in a manner almost identical to host photoreceptors. **d**, Confocal section through inner ONL of recipient retina loaded with calcium indicator. Examples of integrated *Nrl-gfp*^{+/+} (white circles) and recipient photoreceptors (blue circles) are outlined. **e, f**, Intracellular calcium changes evoked by the mGluR8 agonist DCPG are blocked by the specific antagonist CPPG in transplanted and host photoreceptors. Graphs show mean $\pm$ s.e.m.

cone stimulation in $\rho^{-/-}$ mice²⁸. Similarly, we found no measurable response in sham-injected eyes at low light intensities ($N = 4$); threshold responses were observed at $0.052 \text{ cd s}^{-1} \text{ m}^{-2}$ (Fig. 4d, e). By contrast, threshold responses could be elicited in the treated eyes by stimuli as low as $5.7 \times 10^{-3} \text{ cd s}^{-1} \text{ m}^{-2}$ ($N = 4$; Fig. 4d, e), well within the rod photoreceptor range²⁸. In uninjected

wild-type mice, responses were evoked at $4.1 \times 10^{-4} \text{ cd s}^{-1} \text{ m}^{-2}$ ($N = 3$). These results show that integrated cells respond to light and make functional synaptic connections to downstream retinal targets.

Light-induced pupil constriction is a behavioural response that depends on photoreceptors having functional connections with

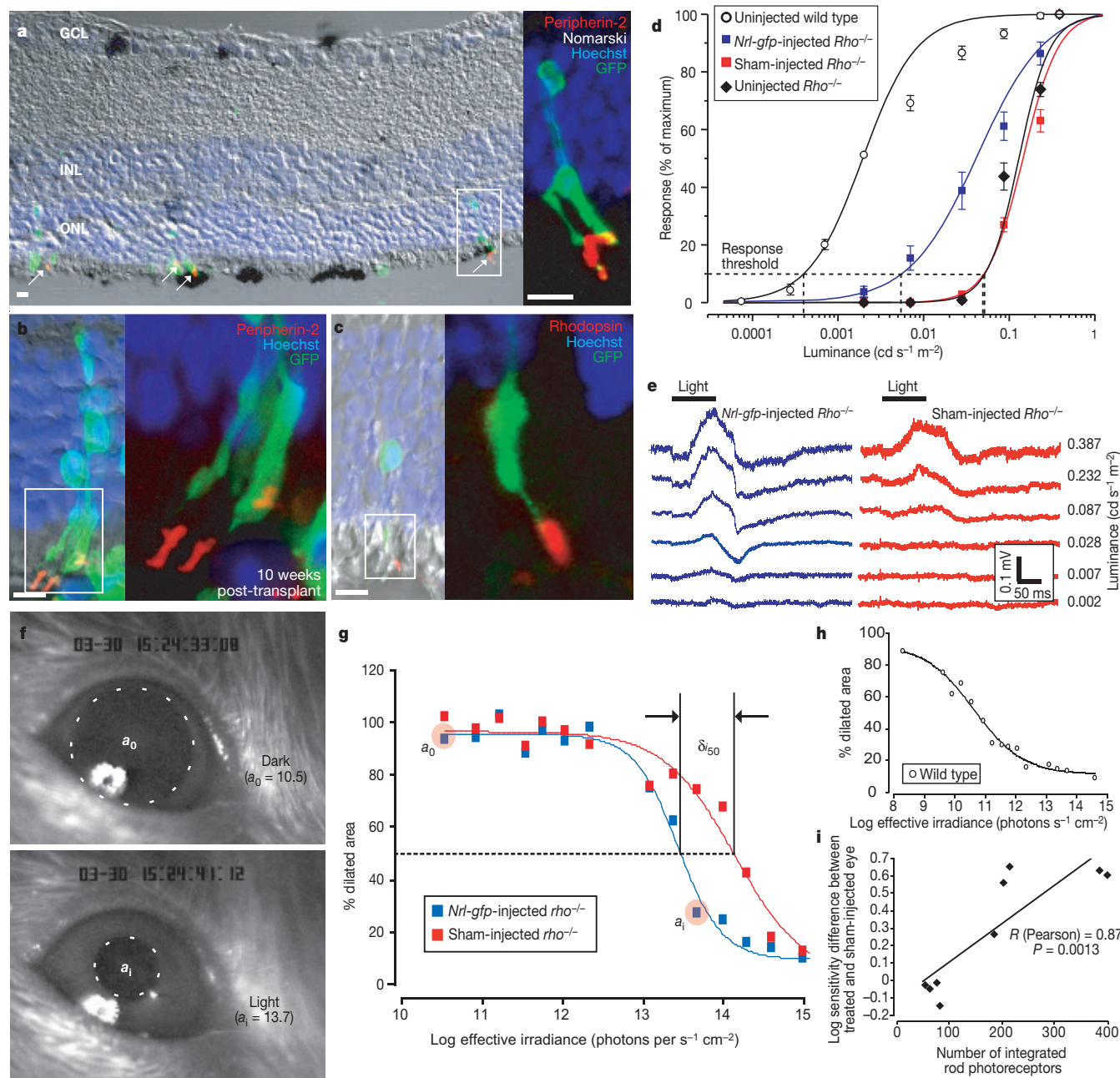


Figure 4 | Integration and restoration of light sensitivity in degenerating recipient retinas. **a, b**, Integration into peripherin-2-deficient $\rho^{-/-}$ mice. **a**, Peripherin-2 (red) in outer segment of GFP-positive cells (arrows). **b**, Peripherin-2 expression is maintained >10 weeks after transplantation. **c**, Left, rhodopsin (red) in outer segment of GFP-positive cells integrated in rhodopsin-deficient $\rho^{-/-}$ mouse. Highlighted sections in **a–c** are shown enlarged (right). Scale bars, 10 μm . **d, e**, Light-evoked extracellular field potentials in the GCL of transplanted retinas. **d**, Light-intensity plots (see Supplementary Information for methods) show that the threshold for response is shifted in treated ($Nrl\text{-}gfp^{+/+}\rho^{+/+}$ cells) (blue squares; $N = 4$) as compared with sham-injected ($\rho^{-/-}$ cells) (red squares; $N = 4$) eyes. Plots for uninjected wild-type (white circles; $N = 3$) and $\rho^{-/-}$ (black diamonds; $N = 3$) eyes are shown for comparison. Plots show means $\pm$ s.d. **e**, Representative recordings from treated (blue) and sham-injected (red)

eyes of the same animal, showing averaged voltage responses to light stimuli of increasing intensity. **f–i**, Light-evoked pupillary responses.

f, Representative infrared images of pupil area measured in dark (a_0) and in light (a_i) corresponding to shaded circles in **g**. **g**, Pupillary response plots ((a_i/a_0) against $\log(i)$, where i is intensity of irradiance) for (**g**) a $\rho^{-/-}$ mouse that received $Nrl\text{-}gfp^{+/+}\rho^{+/+}$ cells (blue squares) in one eye and a sham injection ($\rho^{-/-}$ cells) (red squares) in the other, and (**h**) an uninjected wild-type mouse. Note the increased sensitivity of the $Nrl\text{-}gfp^{+/+}\rho^{+/+}$ -injected eye (blue curve) compared with the sham-injected eye (red curve). **i**, Difference between treated and control in $\log(i)$ required to elicit 50% pupil constriction (δi_{50}), plotted against the number of integrated rod photoreceptors identified histologically. Increasing values on the y axis represent an increase in sensitivity of the treated eye relative to the sham-injected eye.

central brainstem targets. The pupil responses of uninjected wild-type mice ($N=3$) and $\rho^{-/-}$ mice that had received $Nrl\text{-}gfp^{+/+}\rho^{+/+}$ donor cells into one eye and sham injections ($\rho^{-/-}$) into the other ($N=9$) were examined at various intensities (Fig. 4f–i). Wild-type pupils were ~ 3.15 log units more sensitive than the sham-injected eyes of $\rho^{-/-}$ mice (Fig. 4g, h). Sham-injected eyes in $\rho^{-/-}$ mice had no discernible pupil reflex at low light intensities (Fig. 4h). However, in five out of nine $\rho^{-/-}$ mice, the eye that had been injected with $Nrl\text{-}gfp^{+/+}\rho^{+/+}$ cells was more sensitive than the sham-injected eye (Fig. 4h). There was no difference between the two eyes in the remaining four animals. After pupil assessment, the eyes were examined histologically for evidence of cell integration into the ONL. Across all nine animals, the difference in pupil sensitivity compared with the control eye was correlated with the number of $Nrl\text{-}gfp^{+/+}\rho^{+/+}$ cells that had integrated into the ONL (Fig. 4i). These data indicate that integrated cells are light responsive and make functional connections to the brain.

In summary, we have shown that adult wild-type and degenerating mammalian retinas can effectively incorporate rod photoreceptor precursor cells into the ONL. These cells differentiate, form functional synaptic connections with downstream targets in the recipient retina and contribute to visual function. Rather than the environment of the mature retina inhibiting photoreceptor maturation, we show that transplantation of precursor cells at a specific ontogenetic stage, defined by activation of the transcription factor *Nrl*, results in their integration and subsequent differentiation into rod photoreceptors, even in retinal degeneration. Conversely, progenitor or stem cells that have not yet begun to express *Nrl* do not show this property and fail to integrate. Identification of the optimal ontogenetic stage for donor cells might facilitate the generation of appropriate cells for transplantation into humans from either embryonic or adult-derived stem cells.

METHODS

See Supplementary Information for experimental details.

Neural retinas from P1 $Cba\text{-}gfp^{+/+}$ or $Nrl\text{-}gfp^{+/+}$ mice were dissociated enzymatically to a single cell suspension ($\sim 4 \times 10^5$ cells μl^{-1}). Cells were transplanted subretinally into either P1 or adult wild-type mice or mouse models of retinal degeneration. Recipient animals were killed 3 weeks after transplantation, unless otherwise stated. Eyes were fixed and cryosectioned before analysis. Integrated cells were assessed morphologically and immunohistochemically, using confocal microscopy. See Supplementary Information for details of immunohistochemistry, BrdU labelling and calcium imaging protocols. Figures show projection images of 10–15- μm stacks or single confocal sections, where appropriate.

Pupillometry. Pupil reflexes were examined in dark-adapted unanaesthetized animals by using an infrared camera. Animals were subjected to 10-s white light exposures of ascending irradiance. Pupil area was determined 5 s after light exposure (a_i) and expressed relative to the baseline dilated area (a_0).

Extracellular field potential recordings. Recordings were made from the GCL of dark-adapted animals by using glass microelectrodes (1–3 M Ω). Light-evoked potentials were stimulated by 100-ms flashes (green LED, 562-nm peak wavelength) of increasing intensity. Average light-intensity plots were determined from voltage responses evoked by 10–20 flashes of each intensity, from >8 regions of interest. The stimulus threshold was the intensity that evoked a response $>10\%$ of that evoked by the maximum stimulus.

N = number of eyes, n = number of cells examined.

Received 2 June 2006; accepted 10 August 2006.

- Chacko, D. M., Rogers, J. A., Turner, J. E. & Ahmad, I. Survival and differentiation of cultured retinal progenitors transplanted in the subretinal space of the rat. *Biochem. Biophys. Res. Commun.* **268**, 842–846 (2000).
- Sakaguchi, D. S. *et al.* Transplantation of neural progenitor cells into the developing retina of the Brazilian opossum: an *in vivo* system for studying stem/progenitor cell plasticity. *Dev. Neurosci.* **26**, 336–345 (2004).
- Van Hoffelen, S. J., Young, M. J., Shatos, M. A. & Sakaguchi, D. S. Incorporation of murine brain progenitor cells into the developing mammalian retina. *Invest. Ophthalmol. Vis. Sci.* **44**, 426–434 (2003).
- Young, M. J., Ray, J., Whiteley, S. J., Klassen, H. & Gage, F. H. Neuronal differentiation and morphological integration of hippocampal progenitor cells transplanted to the retina of immature and mature dystrophic rats. *Mol. Cell. Neurosci.* **16**, 197–205 (2000).

- Young, R. W. Cell differentiation in the retina of the mouse. *Anat. Rec.* **212**, 199–205 (1985).
- Akimoto, M. *et al.* Targeting of green fluorescent protein to new-born rods by *Nrl* promoter and temporal expression profiling of flow-sorted photoreceptors. *Proc. Natl Acad. Sci. USA* **103**, 3890–3895 (2006).
- Okabe, M., Ikawa, M., Kominami, K., Nakanishi, T. & Nishimune, Y. 'Green mice' as a source of ubiquitous green cells. *FEBS Lett.* **407**, 313–319 (1997).
- Yang, P., Seiler, M. J., Aramant, R. B. & Whittemore, S. R. Differential lineage restriction of rat retinal progenitor cells *in vitro* and *in vivo*. *J. Neurosci. Res.* **69**, 466–476 (2002).
- Carter-Dawson, L. D. & LaVail, M. M. Rods and cones in the mouse retina. II. Autoradiographic analysis of cell generation using tritiated thymidine. *J. Comp. Neurol.* **188**, 263–272 (1979).
- Terada, N. *et al.* Bone marrow cells adopt the phenotype of other cells by spontaneous cell fusion. *Nature* **416**, 542–545 (2002).
- Ying, Q. L., Nichols, J., Evans, E. P. & Smith, A. G. Changing potency by spontaneous fusion. *Nature* **416**, 545–548 (2002).
- Weimann, J. M., Johansson, C. B., Trejo, A. & Blau, H. M. Stable reprogrammed heterokaryons form spontaneously in Purkinje neurons after bone marrow transplant. *Nature Cell Biol.* **5**, 959–966 (2003).
- Hadjantonakis, A. K., Macmaster, S. & Nagy, A. Embryonic stem cells and mice expressing different GFP variants for multiple non-invasive reporter usage within a single animal. *BMC Biotechnol.* **2**, 11 (2002).
- Kashofer, K. & Bonnet, D. Gene therapy progress and prospects: stem cell plasticity. *Gene Ther.* **12**, 1229–1234 (2005).
- Swaroop, A. *et al.* A conserved retina specific gene encodes a basic motif/leucine zipper domain. *Proc. Natl Acad. Sci. USA* **89**, 266–270 (1992).
- Mears, A. J. *et al.* *Nrl* is required for rod photoreceptor development. *Nature Genet.* **29**, 447–452 (2001).
- Swain, P. K. *et al.* Multiple phosphorylated isoforms of *NRL* are expressed in rod photoreceptors. *J. Biol. Chem.* **276**, 36824–36830 (2001).
- tom Dieck, S. *et al.* Molecular dissection of the photoreceptor ribbon synapse: physical interaction of Bassoon and RIBEYE is essential for the assembly of the ribbon complex. *J. Cell Biol.* **168**, 825–836 (2005).
- Koulen, P., Kuhn, R., Wassle, H. & Brandstätter, J. H. Modulation of the intracellular calcium concentration in photoreceptor terminals by a presynaptic metabotropic glutamate receptor. *Proc. Natl Acad. Sci. USA* **96**, 9909–9914 (1999).
- Koulen, P. & Brandstätter, J. H. Pre- and postsynaptic sites of action of mGluR8a in the mammalian retina. *Invest. Ophthalmol. Vis. Sci.* **43**, 1933–1940 (2002).
- Wells, J. *et al.* Mutations in the human retinal degeneration slow (RDS) gene can cause either retinitis pigmentosa or macular dystrophy. *Nature Genet.* **3**, 213–218 (1993).
- McLaughlin, M. E., Sandberg, M. A., Berson, E. L. & Dryja, T. P. Recessive mutations in the gene encoding the beta-subunit of rod phosphodiesterase in patients with retinitis pigmentosa. *Nature Genet.* **4**, 130–134 (1993).
- Rosenfeld, P. J. *et al.* A null mutation in the rhodopsin gene causes rod photoreceptor dysfunction and autosomal recessive retinitis pigmentosa. *Nature Genet.* **1**, 209–213 (1992).
- Reuter, J. H. & Sanyal, S. Development and degeneration of retina in rds mutant mice: the electroretinogram. *Neurosci. Lett.* **48**, 231–237 (1984).
- Sanyal, S., Hawkins, R. K. & Zeilmaker, G. H. Development and degeneration of retina in rds mutant mice: analysis of interphotoreceptor matrix staining in chimeric retina. *Curr. Eye Res.* **7**, 1183–1190 (1988).
- Carter-Dawson, L. D., LaVail, M. M. & Sidman, R. L. Differential effect of the rd mutation on rods and cones in the mouse retina. *Invest. Ophthalmol. Vis. Sci.* **17**, 489–498 (1978).
- Humphries, M. M. *et al.* Retinopathy induced in mice by targeted disruption of the rhodopsin gene. *Nature Genet.* **15**, 216–219 (1997).
- Toda, K., Bush, R. A., Humphries, P. & Sieving, P. A. The electroretinogram of the rhodopsin knockout mouse. *Vis. Neurosci.* **16**, 391–398 (1999).
- Lucas, R. J., Douglas, R. H. & Foster, R. G. Characterization of an ocular photopigment capable of driving pupillary constriction in mice. *Nature Neurosci.* **4**, 621–626 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Y. Duran and N. Gent for technical assistance, P. Humphries for providing the rhodopsin knockout mouse, R. Molday and G. Travis for providing antibodies, and J. Partridge for light calibrations. This work was supported by grants from the Medical Research Council UK, the Royal Blind Asylum and School and The Scottish National Institute for the War Blind. Development of the $Nrl\text{-}gfp^{+/+}$ transgenic line was supported by grants from the National Institutes of Health, The Foundation Fighting Blindness and Research to Prevent Blindness.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.R.A. (r.ali@ucl.ac.uk) or A.S. (swaroop@umich.edu).

LETTERS

Spider toxins activate the capsaicin receptor to produce inflammatory pain

Jan Siemens¹, Sharleen Zhou², Rebecca Piskorowski^{3†}, Tetsuro Nikai⁴, Ellen A. Lumpkin^{3†}, Allan I. Basbaum⁴, David King² & David Julius¹

Bites and stings from venomous creatures can produce pain and inflammation as part of their defensive strategy to ward off predators or competitors^{1,2}. Molecules accounting for lethal effects of venoms have been extensively characterized, but less is known about the mechanisms by which they produce pain. Venoms from spiders, snakes, cone snails or scorpions contain a pharmacopoeia of peptide toxins that block receptor or channel activation as a means of producing shock, paralysis or death^{3–5}. We examined whether these venoms also contain toxins that activate (rather than inhibit) excitatory channels on somatosensory neurons to produce a noxious sensation in mammals. Here we show that venom from a tarantula that is native to the West Indies contains three inhibitor cysteine knot (ICK) peptides that target the capsaicin receptor (TRPV1), an excitatory channel expressed by sensory neurons of the pain pathway⁶. In contrast with the predominant role of ICK toxins as channel inhibitors^{5,7}, these previously unknown ‘vanillotoxins’ function as TRPV1 agonists, providing new tools for understanding mechanisms of TRP channel gating. Some vanillotoxins also inhibit voltage-gated potassium channels, supporting potential similarities between TRP and voltage-gated channel structures. TRP channels can now be included among the targets of peptide toxins, showing that animals, like plants (for example, chilli peppers), avert predators by activating TRP channels on sensory nerve fibres to elicit pain and inflammation.

To identify previously unknown activators of sensory neurons, we examined venoms from 22 spider and scorpion species whose bites are known to produce pain. Screening targets included three members of the TRP ion-channel family whose activation leads to the direct depolarization of primary afferent neurons, consequently producing irritation or pain. These include TRPV1, TRPA1 and TRPM8, which are activated by plant-derived irritants such as capsaicin, mustard oil and menthol, respectively^{8,9}. Using calcium imaging as a functional readout, we found that robust and reproducible signals were obtained when crude venom from *Psalmopoeus cambridgei*, a tarantula native to the West Indies, was applied to TRPV1-expressing HEK-293 cells. In contrast, no activity was observed with cells expressing TRPA1 or TRPM8 (Fig. 1a), nor did the venom inhibit activation of these channels by their cognate chemical agonists (not shown).

Reverse-phase chromatographic methods were used to purify the active component(s) from *P. cambridgei* venom. Three fractions with TRPV1 agonist activity were identified (Supplementary Fig. 1) and presumed to consist of peptides on the basis of retention times and

molecular masses (observed monoisotopic masses were 4,008.5, 3,842 and 4,179 Da, respectively). Indeed, Edman sequencing revealed a family of three closely related peptides, which we have named vanillotoxins (VaTx) 1, 2 and 3. Each peptide consisted of 34 or 35 amino acid residues plus an amidated carboxy terminus (Fig. 1b), with calculated molecular masses (4,008.5, 3,842.5 and 4,179.6 Da, respectively) matching the observed masses noted above.

Vanillotoxins are new members of the extended family of ICK peptides from spiders and cone snails¹⁰. ICK toxins are widely recognized as blockers of cationic channels, perhaps best characterized by the interaction of hanatoxins (HaTx1 and HaTx2) with voltage-gated (Kv) potassium channels^{7,11}. The typical ICK toxin is a 30–40-residue peptide in which three disulphide bridges form a cysteine knot, constraining the molecule into a rigid and compact structure. A short (five to seven residues) hypervariable β -hairpin turn protruding from this compact core is believed to be critical for specifying toxin–ion–channel interactions^{12–14}. Interestingly, VaTx1 and VaTx2 are identical at five of the six residues in this hypervariable turn, whereas VaTx3 shows significant diversity within this region (Fig. 1b). VaTx1 and VaTx2 show greatest identity to heteroscedratxin 1 (HmTx1) (Fig. 1c), an inhibitor of Kv2-type voltage-gated potassium channels¹⁵. VaTx3 is more closely related to the spider toxin, huwentoxin 5 (HwTx-V), from *Ornithoconus huwena*, for which a molecular target is not known¹⁶. Intrigued by the striking similarity between these latter two toxins, we examined whether crude venom from *O. huwena* could activate the capsaicin receptor. Indeed, application of highly diluted (1:4,000) venom to TRPV1-expressing HEK-293 cells produced large calcium responses that were attenuated by ruthenium red, a non-selective blocker of numerous TRP channels (Fig. 1d). In contrast, no response was observed with TRPA1-expressing cells. These results show that *P. cambridgei* is not unique in targeting TRPV1, indicating that VaTx-like peptides might contribute to the pain-producing actions of other spider venoms.

To show unequivocally that VaTx peptides account for TRPV1 activation by *P. cambridgei* venom, we examined whether synthetic VaTx1 peptide could recapitulate the actions of native material. Synthetic VaTx1 was denatured, refolded in glutathione-based redox buffer and subjected to C₈ reverse-phase chromatography, revealing a series of overlapping peaks whose most prominent species had a retention time matching that of native VaTx1 peptide (Supplementary Fig. 2). Refolded synthetic material elicited robust TRPV1 activation, confirming the identity of VaTx1 as an active constituent of *P. cambridgei* venom.

¹Department of Cellular and Molecular Pharmacology, University of California–San Francisco, 600 16th Street, San Francisco, California 94143-2140, USA. ²Howard Hughes Medical Institute Mass Spectrometry Laboratory, University of California–Berkeley, California 94720-3202, USA. ³Department of Physiology, University of California–San Francisco, 600 16th Street, San Francisco, California 94143-2280, USA. ⁴Departments of Anatomy and Physiology and W. M. Keck Center for Integrative Neuroscience, University of California–San Francisco, 1550 4th Street, San Francisco, California 94143-2722, USA. [†]Present addresses: Center for Neurobiology and Behavior, HHMI Columbia University, 1051 Riverside Drive, PI Annex, Room 633 New York, New York 10032, USA (R.P.); Department of Neuroscience, Baylor College of Medicine, One Baylor Plaza, Room 5636-A, Houston, Texas 77030, USA (E.A.L.).

As observed with crude venom, each of the purified vanillotoxins selectively activated HEK-293 cells expressing TRPV1, but not TRPA1 or TRPM8 (Supplementary Fig. 3). Dose-response analysis by calcium imaging revealed a rank order potency of $\text{VaTx3} > \text{VaTx2} > \text{VaTx1}$, with half-maximal effective concentrations (EC_{50}) of 0.45 ± 0.04 , 1.35 ± 0.15 and $9.9 \pm 0.97 \mu\text{M}$ (means $\pm$ s.e.m.), respectively (Fig. 2a). We found that vanillotoxins are present in crude venom at concentrations greatly exceeding their potency at TRPV1, such that toxin concentration and potency are inversely correlated, with VaTx3 being least abundant ($180 \mu\text{M}$) and VaTx1 most prevalent ($1,700 \mu\text{M}$). Because TRPV1 is a heat-sensitive channel⁶, the most physiologically relevant measure of toxin potency may be that assessed at skin temperature, corresponding to the site of venom injection. Under these conditions (34°C), the EC_{50} for VaTx3 was substantially enhanced, to $0.2 \pm 0.03 \mu\text{M}$. Taken together, our findings demonstrate that TRPV1 is a specific and physiologically relevant target for vanillotoxins.

Whole-cell patch-clamp analysis of TRPV1-transfected HEK-293 cells showed that vanillotoxin-evoked currents are characterized by pronounced outward rectification, resembling that observed with capsaicin and other TRPV1 agonists (Fig. 2b, and Supplementary

Fig. 4a). Thermosensitive TRP channels, such as TRPV1 and TRPM8, also show modest voltage dependence, a property that might contribute to gating by chemical or thermal stimuli¹⁷. In this regard, we analysed tail currents from TRPV1-expressing HEK-293 cells (Supplementary Fig. 4b) and found that conductance-voltage relations for capsaicin and VaTx3 were essentially superimposable except for a small but significant difference in slope (0.6 ± 0.0 and 0.9 ± 0.2 for capsaicin and VaTx3, respectively; $P = 0.0017$, Student's t -test) (Fig. 2b). Thus we conclude that these agonists gate TRPV1 by a similar mechanism, although the toxin may have a somewhat greater effect on voltage dependence of gating.

Capsaicin binds to residues on TRPV1 that face the cytoplasm or reside within the inner leaflet of the lipid bilayer^{18–20}. In contrast, ICK toxins have been proposed to interact with their targets by binding to an exposed extracellular loop²¹ or partitioning into the outer leaflet of the plasma membrane to contact residues within the bilayer^{11,22,23}. It

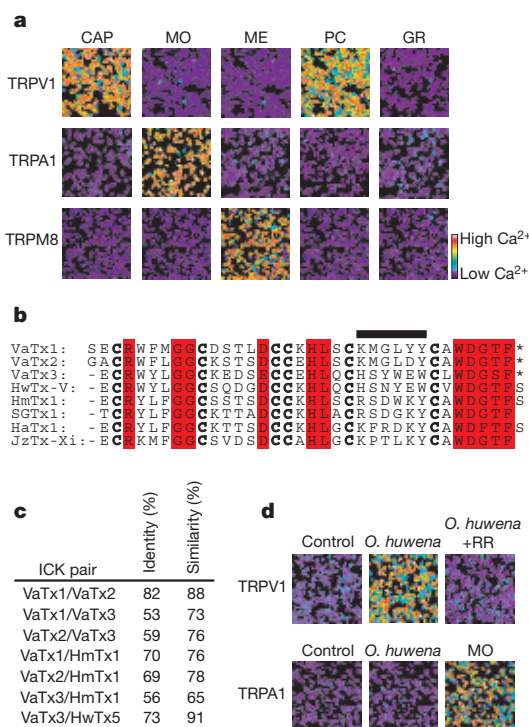


Figure 1 | A subfamily of peptides from tarantula venom activates TRPV1. **a**, HEK-293 cells expressing TRPV1, TRPM8 or TRPA1 were challenged with capsaicin (CAP; 40 nM), allyl isothiocyanate (mustard oil, MO; 100 μM), menthol (ME; 75 μM) or crude *P. cambridgei* (PC) or *Grammostola rosae* (GR) venom (diluted 1:1,000), as indicated. Channel activation was assessed by ratiometric calcium imaging (the colour bar indicates relative calcium levels). Similar results were obtained with nine independent venom lots ($n \geq 400$ cells per visual field). **b**, Alignment of *P. cambridgei* vanillotoxins (VaTx1, VaTx2 and VaTx3) with related ICK peptides from *O. huwena* (HwTx-V), *H. maculata* (HmTx1), *Scodra griseipes* (SGTx1), *Grammostola spatulata* (HaTx1) and *Chilobrachys jingzhao* (JzTx-Xi) spiders. Cysteine residues characteristic of ICK peptides are shown in bold and other highly conserved residues are highlighted in red. The black bar indicates the hypervariable turn. Asterisks denote C-terminal amidation in vanillotoxins. **c**, Pairwise comparison of ICK peptides showing percentage amino acid identities and similarities. **d**, Venom from *O. huwena* (1:4,000) elicited calcium influx into HEK-293 cells expressing TRPV1 (top) that was blocked by ruthenium red (+RR; 10 μM). Cells expressing TRPA1 were not affected by *O. huwena* venom but responded to mustard oil (MO; 100 μM) ($n \geq 400$ cells per visual field).

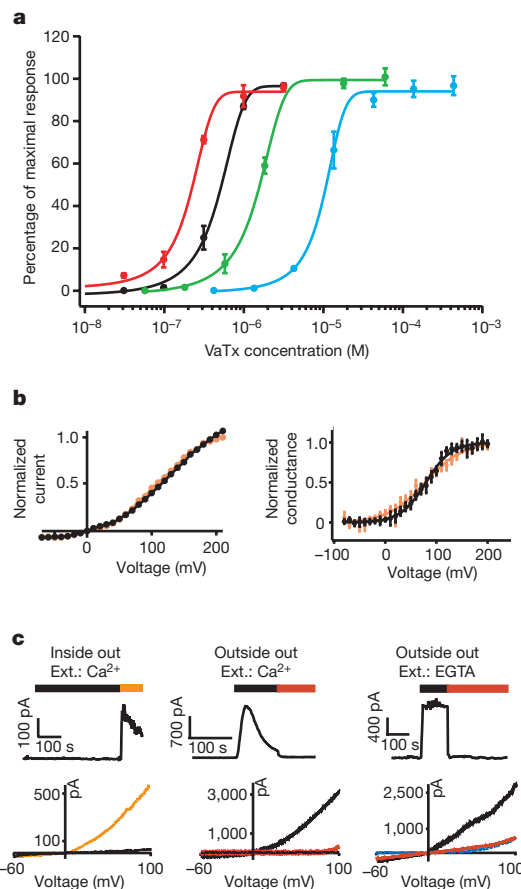


Figure 2 | Functional properties of purified vanillotoxins. **a**, Dose-response analysis of vanillotoxins (blue, VaTx1; green, VaTx2; black, VaTx3) by ratiometric calcium imaging of TRPV1-transfected HEK-293 cells. Responses at each toxin concentration were determined at room temperature (24°C) and normalized to the response at saturating capsaicin concentration (5 μM). VaTx3 potency was also determined at 34°C (red) ($n = 3$ trials per concentration). **b**, Currents evoked by capsaicin (0.2 μM ; orange) and VaTx3 (5 μM ; black) were recorded from TRPV1-transfected HEK-293 cells in whole-cell configuration, from which current-voltage (left) and conductance-voltage (right) relationships were derived. Error bars represent s.e.m.; $n = 4$ for VaTx3 and $n = 5$ for capsaicin. Plots were generated from current traces shown in Supplementary Fig. 4. **c**, Inside-out (left) and outside-out (middle and right) patches were excised from TRPV1-expressing HEK-293 cells, exposed to VaTx3 (black bar, 0.5 μM), capsaicin (orange bar, 2 μM) or ruthenium red (red bar, 10 μM) as indicated, and currents were recorded at +60 mV in calcium (1.5 mM)-containing or calcium-free (1.0 mM EGTA) bath solution. Current-voltage relations are shown below each trace (blue curve indicates no agonist control) ($n \geq 5$ patches per configuration).

therefore seems likely that vanillotoxins and capsaicin interact with TRPV1 at opposite faces of the membrane. To address this question, we examined whether VaTx3 could activate channels when applied to inside-out versus outside-out membrane patches excised from TRPV1-expressing HEK-293 cells. Indeed, the toxin elicited responses in outside-out patches only, whereas capsaicin was effective in either configuration (Fig. 2c, and not shown). Parenthetically, VaTx3-evoked currents showed calcium-dependent desensitization (Fig. 2c) resembling that observed with capsaicin²⁴. Taken together, our observations indicate that capsaicin and vanillotoxins share similarities in their mechanisms of TRPV1 activation, but probably interact with distinct regions of the channel. Consistent with this was our observation that vanillotoxins activate avian TRPV1 channels (not shown), which are insensitive to capsaicin and structurally related agonists¹⁹.

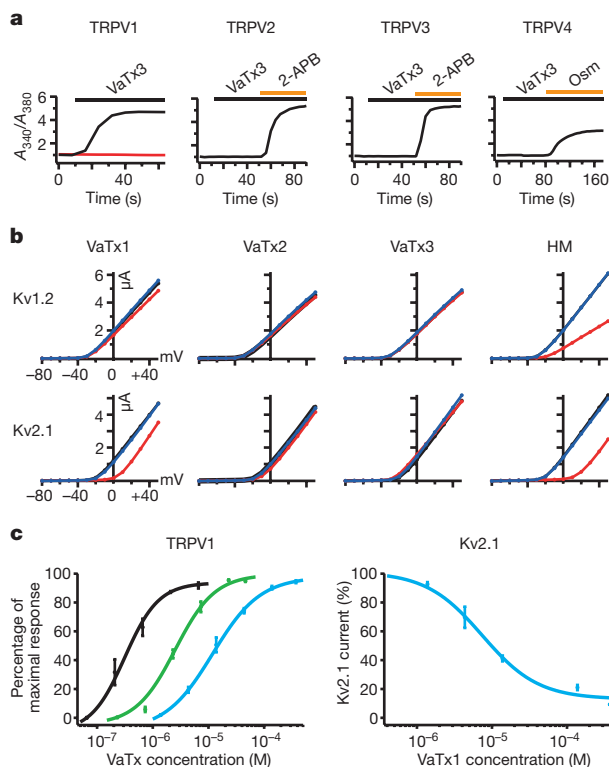


Figure 3 | Analysis of vanillotoxin specificity. **a**, VaTx3 activates TRPV1 but not the related thermosensitive or osmosensitive channels TRPV2, TRPV3 or TRPV4. HEK-293 cells expressing indicated channels were exposed to VaTx3 (0.5 μ M) and responses were analysed by ratiometric calcium imaging. To control for functional expression, TRPV2, TRPV3 and TRPV4 were subsequently challenged with the broad-spectrum agonist 2-aminoethyl diphenylborinate (2-APB; 250 μ M), or hypo-osmotic Ringer solution (Osm; 200 mOsm), as indicated. The red trace shows blockade of toxin response by ruthenium red (10 μ M). Traces represent an average of $n \geq 80$ cells. **b**, Currents were elicited in oocytes expressing voltage-gated potassium channels, Kv1.2 or Kv2.1, by applying depolarizing pulses (-100 to $+50$ mV) from a -80 mV holding potential in 10-mV steps (each lasting 500 ms). Current-voltage relationships were plotted before the addition (black), after the addition (red) and after the wash-out (blue) of the respective toxin (VaTx1, 20 μ M; VaTx2, 5 μ M; VaTx3, 0.75 μ M). **H. maculata** (HM) venom (dilution 1:200) served as positive control for Kv2.1 and Kv1.2 inhibition ($n = 4$ independent experiments per toxin). **c**, Left: dose-response analysis by whole-cell voltage-clamp recording of TRPV1-expressing oocytes ($+80$ mV) revealed EC_{50} values of 11.9 ± 1.4 , 2.53 ± 0.02 and 0.32 ± 0.09 μ M for VaTx1 (blue), VaTx2 (green) and VaTx3 (black), respectively. All values are in excellent agreement with those obtained by calcium imaging (see Fig. 2a). Right: toxin-mediated inhibition of Kv2.1 (at 0 mV) was assessed in oocytes. The full dose-response is shown for VaTx1 ($IC_{50} = 7.4 \pm 1.9$ μ M). Error bars represent s.e.m.; $n \geq 3$ trials for each toxin concentration.

To address vanillotoxin specificity further, we examined their effects on other TRPV family members, including TRPV2, TRPV3 and TRPV4. Neither VaTx1, VaTx2 nor VaTx3 activated HEK-293 cells expressing these channels (Fig. 3a, and not shown). In light of the well-established interactions between ICK peptides (for example hanatoxins and heteroscodratoxins) and voltage-gated potassium channels, we also examined whether vanillotoxins modulate the activity of Kv1.2, Kv2.1 or Kv4.2 channels, representing Shaker, Shab and Shal subfamilies, respectively. Each of these Kv channels was expressed in oocytes and gating in response to voltage steps (-100 to $+50$ mV) assessed in the absence or presence of purified vanillotoxins. Significant inhibition of Kv2.1 was observed with VaTx1 (Fig. 3b, and Supplementary Fig. 5). Moreover, tail current analysis revealed that VaTx1 shifted Kv2.1 activation to higher voltages (Supplementary Fig. 5), resembling the inhibitory effects of hanatoxins and heteroscodratoxins^{11,15}. Detailed dose-response analysis showed that VaTx1 is equally potent as a TRPV1 agonist or a Kv2.1 antagonist ($EC_{50} = 11.9 \pm 1.4$ μ M and $IC_{50} = 7.4 \pm 1.9$ μ M, respectively) (Fig. 3c). VaTx2 also inhibited Kv2.1, but only at concentrations severalfold exceeding its EC_{50} at TRPV1. VaTx3 showed greatest selectivity for TRPV1, producing little or no inhibition of Kv channels at concentrations 10–100-fold exceeding its EC_{50} at TRPV1 (Supplementary Fig. 5). Conversely, *Heteroscodrata maculata* venom neither activated nor inhibited TRPV1 (Supplementary Fig. 6), demonstrating specificity of these ICK toxins for their respective channel targets.

To assess vanillotoxin selectivity *in vivo*, we first examined their actions on dissociated sensory neurons from mouse trigeminal ganglia. In cultures from wild-type mice, VaTx2 or VaTx3 produced robust calcium increases among the subpopulation of capsaicin-sensitive cells (Fig. 4a, b, and not shown), whereas neurons from TRPV1-deficient mice were completely unresponsive. Moreover, even crude venom (at 1:500 dilution) produced rapid and robust calcium increases in capsaicin-sensitive neurons (Supplementary Fig. 7). A significantly smaller response was observed in all other neurons, as well as in non-excitable cells such as fibroblasts, presumably reflecting the effects of other peptides, proteases, lipases, and so on, present in crude venom²⁵. Consistent with this was our observation that neurons from TRPV1-deficient mice showed only this smaller 'background' calcium response (Supplementary Fig. 7). Together with our biochemical analysis, these results show that crude *P. cambridgei* venom contains ample vanillotoxin to activate sensory neurons and that most (but not all) of this response is mediated through the activation of TRPV1.

In wild-type mice, injection of capsaicin into the hind paw elicits pain-related behaviours, such as licking and flinching of the affected limb. Injection of purified VaTx3 produced similar responses in wild-type mice but not in TRPV1-deficient littermates (Fig. 4c), further demonstrating specificity of the toxin for the capsaicin receptor *in vivo*. We also measured paw thickness as an indication of neurogenic inflammation resulting from robust activation of sensory nerve fibres. Indeed, the hind paws of toxin-injected wild-type mice showed substantial oedema, whereas TRPV1-deficient animals displayed only minimal swelling akin to that observed with vehicle-injected wild-type controls (Fig. 4d). Injection of crude venom into the paw produced equivalent behaviour and swelling that was independent of genotype, but this was expected in light of our cellular studies showing a small but finite TRPV1-independent effect of crude venom on sensory neurons. Taken together, these data demonstrate that VaTx peptides are agonists for recombinant or native TRPV1, eliciting pain and inflammation through activation of this excitatory channel.

ICK toxins from spiders, scorpions and predatory cone snails target a variety of cationic channels that are gated by membrane voltage or extracellular ligands. Our findings show that TRP channels can now be included as molecular targets of ICK toxins, or, for that matter, any peptide toxin. With the exception of two scorpion ICK peptides that activate intracellular ryanodine receptors, all other

known toxins of the ICK group function as receptor or channel antagonists¹⁰. Thus, vanillotoxins are unique in serving as TRPV1 agonists, a property that enables them to excite sensory nerve endings to produce pain and inflammation, as commonly associated with bites or stings from venomous creatures. Our results with *O. huwena* venom indicate that a variety of spider species might use ICK toxins to target sensory TRP channels as a strategy to ward off predators or competitors. Indeed, this would parallel the mechanisms adopted by numerous plant species to deter predatory mammals through the production of chemical irritants (such as capsaicin or isothiocyanates) that also target TRP channels on sensory neurons of the pain pathway.

The sequence similarity between vanillotoxins and heteroscodratins or hanatoxins is consistent with the idea that TRP channels resemble voltage-gated potassium channels in their overall transmembrane topology and tetrameric subunit organization²⁶. VaTx1

and VaTx2, which produce significant inhibition of Kv2.1, show greater sequence similarity to each other or to HmTx1 than they do to VaTx3 (Fig. 1b, c), presumably reflecting their relative activities as Kv antagonists and/or TRPV1 agonists. Biochemical and mutagenesis studies indicate that inhibition of Kv channels by hanatoxin might involve interaction of the toxin with regions of the third and fourth transmembrane helices that constitute the voltage sensor domain^{11,27}. Perhaps vanillotoxins interact with a region of TRPV1 that is topologically equivalent to the voltage sensor domain in Kv channels, which may be similarly important in modulating the transition between open and closed states of the TRPV1 channel complex. Although TRP channels display only modest voltage-dependent behaviour in comparison with Kv channels^{9,17}, our conductance–voltage analysis indicates that vanillotoxins, like capsaicin, exert some effect on the voltage dependence of gating. Future biochemical and structure–function studies will be required to delineate the exact nature of vanillotoxin–channel interactions and determine how this alters the voltage sensitivity of TRP or Kv channels. Vanillotoxins provide new pharmacological tools for addressing these and other questions relating to TRP channel structure and gating mechanisms.

METHODS

Cell lines and plasmid constructs. HEK-293 cells stably expressing rat TRPV1, human TRPA1 or rat TRPM8 were generated with inducible (pCDNA4/TO) or non-inducible (pcDNA3) vectors (Invitrogen). Human TRPV1, rat TRPV2 and human TRPV3 were transiently expressed in HEK-293 cells with the use of Lipofectamine 2000 (Invitrogen). Kv1.2, Kv4.2 and Kv2.1 plasmids were linearized and transcribed *in vitro* with T7 or T3 polymerase (Ambion).

Toxin purification and peptide chemistry. Crude spider venom was obtained from Spider Pharm. Detailed protocols describing vanillotoxin purification and biochemical characterization are provided in Supplementary Fig. 1. Protocols for solid-phase synthesis and refolding of VaTx1 are provided in Supplementary Fig. 2.

Calcium imaging and electrophysiology. Culturing and imaging of sensory neurons was performed as described²⁸. Imaging of HEK-293 cells was performed in 10- μ l adhesive silicone isolators (Invitrogen) attached to poly-(D-lysine)-coated microscope slides (Fisher). Fluorescent images were acquired with Metafluor Software (Molecular Devices) and analysed with automated routines written in Igor Pro (Wavemetrics). All experiments were performed at room temperature (24 °C) except dose–response analysis of VaTx3, which was performed at 34 °C.

Patch-clamp recordings in TRPV1-expressing HEK-293 cells were performed in whole-cell mode or excised-patch configurations as described previously²⁹. Currents were recorded with an Axon 200B amplifier and a Digidata 1321A interface and acquired with pClamp software (Axon Instruments). Pipette resistance ranged from 0.9 to 1.5 M Ω . For excised patch experiments, no series resistance compensation was performed, because series resistance was generally less than 3 M Ω . Signals were filtered at 5 kHz and digitized at 25 μ s. Standard bath and pipette solutions contained (in mM) 140 NaCl, 5 KCl, 2 MgCl₂, 2 EGTA, 10 HEPES, and 10 D-glucose (adjusted to pH 7.4 with NaOH). For the whole-cell tail current analysis, conductance–voltage curves were normalized to the maximal conductance at 200 mV and fitted with a Boltzmann equation ($G = G_{\max}/(1 + \exp[-zF(V - V_{1/2})/RT])$). Oocytes were prepared and subjected to two-electrode voltage-clamp analysis as described previously¹⁹. Cells were bathed in normal ND96 solution for the analysis of Kv channels, or in calcium-free solution (120 mM CsCl, 1 mM EGTA, 10 mM HEPES, 2 mM MgCl₂, pH 7.4) for the analysis of TRPV1 channels.

Behaviour. Mice (20–30 g) were housed with a 12 h light/12 h dark cycle at 21 °C, and experiments were performed under the policies of the International Association for the Study of Pain and UCSF Animal Care and Use Committee. Nocifensive licking responses were measured for a period of 5 min after injection of vehicle or toxin (20 μ l), as described³⁰. Hindpaw thickness was determined with a spring-loaded caliper before and 30 min after injection. Experiments were performed with TRPV1^{+/+} and TRPV1^{-/-} littermates while blind to genotype. Data were analysed with ANOVA single-factor statistical analysis for pairwise comparisons.

Received 26 July; accepted 27 September 2006.

- Schmidt, J. O. Biochemistry of insect venoms. *Annu. Rev. Entomol.* 27, 339–368 (1982).
- Chahl, L. A. & Kirk, E. J. Toxins which produce pain. *Pain* 1, 3–49 (1975).

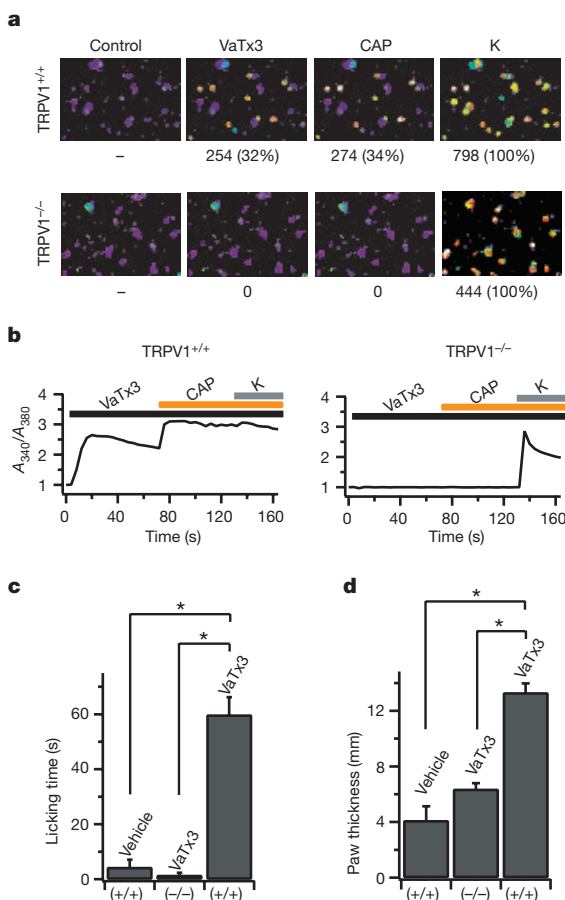


Figure 4 | VaTx3 activates capsaicin-sensitive trigeminal neurons and elicits TRPV1-dependent pain responses in mice. **a**, Trigeminal neurons from wild-type (TRPV1^{+/+}) or TRPV1-deficient (TRPV1^{-/-}) mice were challenged with VaTx3 (1 μ M), capsaicin (CAP; 3 μ M), then high-potassium buffer (K; 100 mM), as indicated, and analysed by ratiometric calcium imaging. The total number of responsive neurons (percentage of excitable cells) is shown below each image. **b**, Average calcium responses for wild-type ($n = 20$) and TRPV1-deficient ($n = 40$) neurons shown in **a**. VaTx3-evoked peak ratiometric responses were 2.6 ± 0.05 and 1.0 ± 0.02 , respectively. **c**, **d**, Hind paws of wild-type (+/+) or TRPV1-deficient (-/-) littermates were injected with VaTx3 (40 μ M in 20 μ l of PBS with 0.5% dimethylsulphoxide) or vehicle alone. Total time spent licking the injected paw was recorded over 5 min (**c**) and change in paw thickness was measured before and 30 min after injection of toxin or vehicle (**d**). Error bars represent s.e.m. ($n = 5$ animals per trial; asterisk, $P \leq 0.0003$, ANOVA single-factor analysis). No significant differences were observed between vehicle-injected TRPV1^{+/+} and toxin-injected TRPV1^{-/-} mice. All measurements were performed blind to genotype.

3. Escoubas, P. & Rash, L. Tarantulas: eight-legged pharmacists and combinatorial chemists. *Toxicon* **43**, 555–574 (2004).
4. Miller, C. The charybdotoxin family of K⁺ channel-blocking peptides. *Neuron* **15**, 5–10 (1995).
5. Terlau, H. & Olivera, B. M. *Conus* venoms: a rich source of novel ion channel-targeted peptides. *Physiol. Rev.* **84**, 41–68 (2004).
6. Caterina, M. J. *et al.* The capsaicin receptor: a heat-activated ion channel in the pain pathway. *Nature* **389**, 816–824 (1997).
7. Swartz, K. J. & MacKinnon, R. An inhibitor of the Kv2.1 potassium channel isolated from the venom of a Chilean tarantula. *Neuron* **15**, 941–949 (1995).
8. Jordt, S. E., McKemy, D. D. & Julius, D. Lessons from peppers and peppermint: the molecular logic of thermosensation. *Curr. Opin. Neurobiol.* **13**, 487–492 (2003).
9. Voets, T., Talavera, K., Owsianik, G. & Nilius, B. Sensing with TRP channels. *Nature Chem. Biol.* **1**, 85–92 (2005).
10. Zhu, S., Darbon, H., Dyason, K., Verdonck, F. & Tytgat, J. Evolutionary origin of inhibitor cystine knot peptides. *FASEB J.* **17**, 1765–1767 (2003).
11. Swartz, K. J. & MacKinnon, R. Mapping the receptor site for hanatoxin, a gating modifier of voltage-dependent K⁺ channels. *Neuron* **18**, 675–682 (1997).
12. Oswald, R. E., Suchyna, T. M., McFeeters, R., Gottlieb, P. & Sachs, F. Solution structure of peptide toxins that block mechanosensitive ion channels. *J. Biol. Chem.* **277**, 34443–34450 (2002).
13. Takahashi, H. *et al.* Solution structure of hanatoxin1, a gating modifier of voltage-dependent K⁺ channels: common surface features of gating modifier toxins. *J. Mol. Biol.* **297**, 771–780 (2000).
14. Wang, J. M. *et al.* Molecular surface of tarantula toxins interacting with voltage sensors in K_v channels. *J. Gen. Physiol.* **123**, 455–467 (2004).
15. Escoubas, P., Diochot, S., Celerier, M. L., Nakajima, T. & Lazdunski, M. Novel tarantula toxins for subtypes of voltage-dependent potassium channels in the Kv2 and Kv4 subfamilies. *Mol. Pharmacol.* **62**, 48–57 (2002).
16. Zhang, P. F., Chen, P., Hu, W. J. & Liang, S. P. Huwentoxin-V, a novel insecticidal peptide toxin from the spider *Selenocosmia huwena*, and a natural mutant of the toxin: indicates the key amino acid residues related to the biological activity. *Toxicon* **42**, 15–20 (2003).
17. Voets, T. *et al.* The principle of temperature-dependent gating in cold- and heat-sensitive TRP channels. *Nature* **430**, 748–754 (2004).
18. Jung, J. *et al.* Capsaicin binds to the intracellular domain of the capsaicin-activated ion channel. *J. Neurosci.* **19**, 529–538 (1999).
19. Jordt, S. E. & Julius, D. Molecular basis for species-specific sensitivity to 'hot' chili peppers. *Cell* **108**, 421–430 (2002).
20. Gavva, N. R. *et al.* Molecular determinants of vanilloid sensitivity in TRPV1. *J. Biol. Chem.* **279**, 20283–20295 (2004).
21. Salinas, M. *et al.* The receptor site of the spider toxin PcTx1 on the proton-gated cation channel ASIC1a. *J. Physiol. (Lond.)* **570**, 339–354 (2006).
22. Lee, S. Y. & MacKinnon, R. A membrane-access mechanism of ion channel inhibition by voltage sensor toxins from spider venom. *Nature* **430**, 232–235 (2004).
23. Suchyna, T. M. *et al.* Bilayer-dependent inhibition of mechanosensitive channels by neuroactive peptide enantiomers. *Nature* **430**, 235–240 (2004).
24. Tominaga, M. *et al.* The cloned capsaicin receptor integrates multiple pain-producing stimuli. *Neuron* **21**, 531–543 (1998).
25. Ramu, Y., Xu, Y. & Lu, Z. Enzymatic activation of voltage-gated potassium channels. *Nature* **442**, 696–699 (2006).
26. Ramsey, I. S., Delling, M. & Clapham, D. E. An introduction to TRP channels. *Annu. Rev. Physiol.* **68**, 619–647 (2006).
27. Phillips, L. R. *et al.* Voltage-sensor activation with a tarantula toxin as cargo. *Nature* **436**, 857–860 (2005).
28. Jordt, S. E. *et al.* Mustard oils and cannabinoids excite sensory nerve fibres through the TRP channel ANKTM1. *Nature* **427**, 260–265 (2004).
29. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Pflügers Arch.* **391**, 85–100 (1981).
30. Martinez-Caro, L. & Laird, J. M. Allodynia and hyperalgesia evoked by sciatic mononeuropathy in NK1 receptor knockout mice. *Neuroreport* **11**, 1213–1217 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Minor and members of his laboratory for Kv4.2 and Kchip complementary RNAs and advice with chromatographic methods; D. Clapham and R. Aldrich for providing TRPV3 and Kv2.1 plasmids, respectively; K. Shokat and J. Blethrow for initial help with mass spectrometry; members of the Julius laboratory for discussions; and R. Nicoll, R. Edwards and H. Chuang for critical reading of the manuscript. This work was supported by NIH grants (to D.J., A.I.B. and E.A.L.) and by postdoctoral fellowships from the Swiss National Science Foundation, Novartis Stiftung, and the International Human Frontier Science Program Organization (to J.S.).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.J. (julius@cmp.ucsf.edu).

Structure of C3b reveals conformational changes that underlie complement activity

Bert J. C. Janssen¹, Agni Christodoulidou², Andrew McCarthy³, John D. Lambris² & Piet Gros¹

Resistance to infection and clearance of cell debris in mammals depend on the activation of the complement system, which is an important component of innate and adaptive immunity^{1,2}. Central to the complement system is the activated form of C3, called C3b, which attaches covalently to target surfaces³ to amplify complement response, label cells for phagocytosis and stimulate the adaptive immune response. C3b consists of 1,560 amino-acid residues and has 12 domains. It binds various proteins and receptors to effect its functions⁴. However, it is not known how C3 changes its conformation into C3b and thereby exposes its many binding sites. Here we present the crystal structure at 4-Å resolution of the activated complement protein C3b and describe the conformational rearrangements of the 12 domains that take place upon proteolytic activation. In the activated form the thioester is fully exposed for covalent attachment to target surfaces and is more than 85 Å away from the buried site in native C3 (ref. 5). Marked domain rearrangements in the α -chain present an altered molecular surface, exposing hidden and cryptic sites that are consistent with known putative binding sites of factor B and several complement regulators. The structural data indicate that the large conformational changes in the proteolytic activation and regulation of C3 take place mainly in the first conversion step, from C3 to C3b. These insights are important for the development of strategies to treat immune disorders that involve complement-mediated inflammation.

The complement system is important for host defence in metazoans¹. In mammals, complement provides an essential link between innate and adaptive immunity². The complement system has also been proposed to be involved in tissue regeneration, neuron development, wound healing and reproduction⁶, indicating that it has a broad role in cell differentiation. The system consists of about 35 cell-surface and soluble plasma proteins. Cleavage of C3 (186 kDa) into C3b (177 kDa) and C3a (9 kDa) is the central step in the complement activation cascade, which can be initiated by three different pathways—the classical, lectin and alternative pathways. C3b covalently attaches to pathogenic, or apoptotic, target surfaces through its reactive thioester moiety⁷ and thereby induces several biological processes. C3b provides a molecular platform for the formation of convertase complexes. Binding of pro-enzyme factor B to C3b and subsequent cleavage of factor B by factor D yields the short-lived ($t_{1/2} \approx 90$ s) C3bBb complex⁸, which converts C3 into C3b and C3a, thereby amplifying the complement response and forming the C3b₂Bb complex that cleaves C5 to initiate the formation of large, membrane-inserted lytic pores. In addition, deposition of C3b on target surfaces generates two important cellular responses. C3b and its proteolytic fragments iC3b and C3dg (Fig. 1) are recognized by complement receptors (CR) on phagocytic cells (for example, CR1 (CD35), CR3 (CD11b/CD18), CR4 (CD11c/CD18)² and CRIg⁹) to induce the

uptake of opsonized particles, and by CR2 (CD35) to upregulate the B-cell response 10,000-fold through the co-stimulatory and B-cell receptor complexes¹⁰. The formation of C3b is therefore a crucial step that requires tight regulation. Host cells express several cell-surface and soluble regulators that disrupt the C3bBb complex or aid in the proteolytic degradation of C3b into iC3b and ultimately to C3dg and C3c^{11,12}. Bacteria and viruses use proteins with similar activities to evade the complement response. Because C3b binds many soluble proteins and cell-surface receptors⁴, whereas C3 has few binding partners, the proteolytic activation of C3 into C3b probably induces conformational changes that form the various cryptic binding sites. Conformation-dependent mechanisms are probably common to the homologous proteins of the C3/ α 2-macroglobulin family, which are an ancient family of host defence proteins.

Here we present the crystal structure of human C3b solved at 4-Å resolution by molecular replacement using the structures of C3, C3c⁵ and C3d¹³; see Methods and Supplementary Table 1 and Supplementary Fig. 1. C3b is generated from C3 by proteolytic removal of the small anaphylatoxin domain (residues 650–726) to produce a β -chain (residues 1–645) and a shortened α -chain, denoted α' (residues 727–1,641). The structure of C3b reveals that this conversion is accompanied by marked structural rearrangements that result in differences in atomic position of up to 95 Å (Fig. 1). The domain rearrangements can be described by a metaphor of a puppeteer with a body formed by the β -ring domains (macroglobulin (MG) 1–6 domains and linker (LNK) domain), shoulders (domains MG7 and MG8), a neck (anchor region), a head (C345C domain) and an arm ('complement C1r/C1s, Uegf, Bmp1' (CUB) domain) with a puppet (thioester-containing domain (TED)). In C3 the puppet (TED domain) is embraced by the arm (CUB) and held against the shoulder (MG8). In C3b the shoulders (MG7 and MG8), neck (anchor) and head (C345C) have twisted, the arm (CUB) has extended downwards and the puppet (TED) has dropped (see Supplementary Table 2 and Movie 1). In its new position, the TED domain contacts MG1 of the β -ring. The single amino-acid residue difference, Arg80Gly in MG1, that is responsible for the C3S/C3F phenotype difference¹⁴ is located in the MG1–TED interface. The overall arrangement of the core domains (MG1–8, LNK, anchor and C345C) resembles more the structural arrangement of the final proteolytic, downregulated fragment C3c than the native C3 molecule. This includes the relocation of the α' amino-terminal segment (α' NT) and the conformational transition of the MG8 $\beta\alpha\beta$ -motif in C3 to a $\beta\alpha\alpha$ -motif in C3b and C3c. The TED domain (which is absent in C3c) takes on the conformation that is seen in the C3d fragment¹³. Thus, in the proteolytic conversion of C3 to C3b, iC3b and finally C3c and C3dg, the main conformational rearrangements occur upon the first activation step that converts the native, resting C3 into the active C3b.

¹Crystal and Structural Chemistry, Bijvoet Center for Biomolecular Research, Faculty of Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands. ²Department of Pathology and Laboratory Medicine, School of Medicine, University of Pennsylvania, Philadelphia 19104, USA. ³European Molecular Biology Laboratory, Grenoble Outstation, 38042 Grenoble, Cedex 9, France.

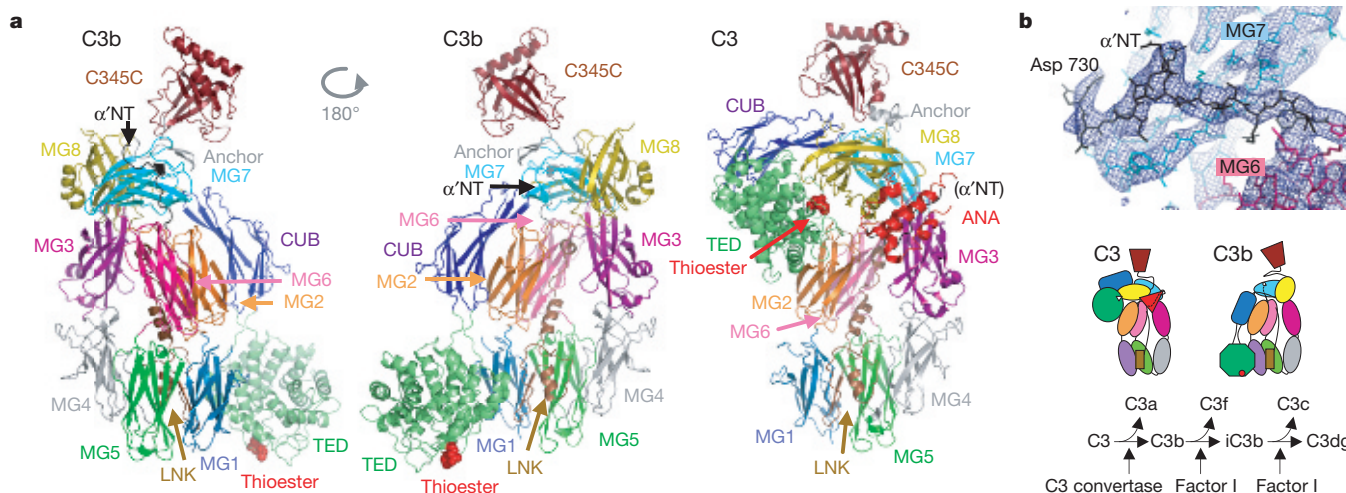


Figure 1 | Structure of C3b at 4-Å resolution. **a**, Ribbon representation in two views of C3b coloured by domain and labelled accordingly. Also indicated are the exposed thioester moiety (red spheres), anchor region (grey) and α' NT (black). For comparison, the ribbon representation of C3 and cartoons of the domain arrangements in C3b and C3 are added. **b**, Electron density ($2mF_{\text{obs}} - DF_{\text{calc}}$) contoured at 1σ of the α' NT region with domains as indicated; density maps for each domain are given in Supplementary Fig. 1. The proteolytic steps of C3 conversion are shown schematically.

Surface attachment of C3b requires exposure and activation of the thioester moiety, Cys 988–Gln 991 of the TED domain. In the structure of C3b these residues are completely exposed to the solvent, whereas in C3 they are hidden⁵ 85 Å away from the site of exposure. In addition, the TED domain in C3b has changed its conformation from C3-like⁵ to C3d-like¹³, where the arrangement of the reactive residues corresponds to the formation of the acyl-imidazole intermediate, which is highly reactive towards hydroxyl nucleophiles^{5,15} (Fig. 2a). The overall domain rearrangements might be coupled to the conformational changes in the TED domain by a pulling force that stretches the N-terminal segment (helices $\alpha 0$ and $\alpha 1$) preceding the thioester cysteine. Surface-charge calculation shows a single, strong electro-positive patch near the thioester, on a protein that is

otherwise mostly and strongly electro-negative (Fig. 2b). Putatively, the electropositive patch on the TED domain is involved in orienting the molecule with respect to negatively charged target surfaces, such as bacterial cells or apoptotic host cells. This orientation results in a footprint of $50 \times 90 \text{ Å}^2$ formed by the TED, MG1, MG4 and MG5 domains, which is in remarkable agreement with the figure of $\sim 4,750 \text{ Å}^2$ that was estimated from fully covered erythrocytes¹⁶. Proper orientation of the TED domain close to the target surface might be important for the highly reactive and short-lived ($t_{1/2} < 100 \mu\text{s}$)¹⁷ acyl-imidazole intermediate to react covalently with the target surface instead of the surrounding solvent.

Conversion of C3 to C3b exposes binding sites for proteins involved in convertase formation (factor B and properdin) and

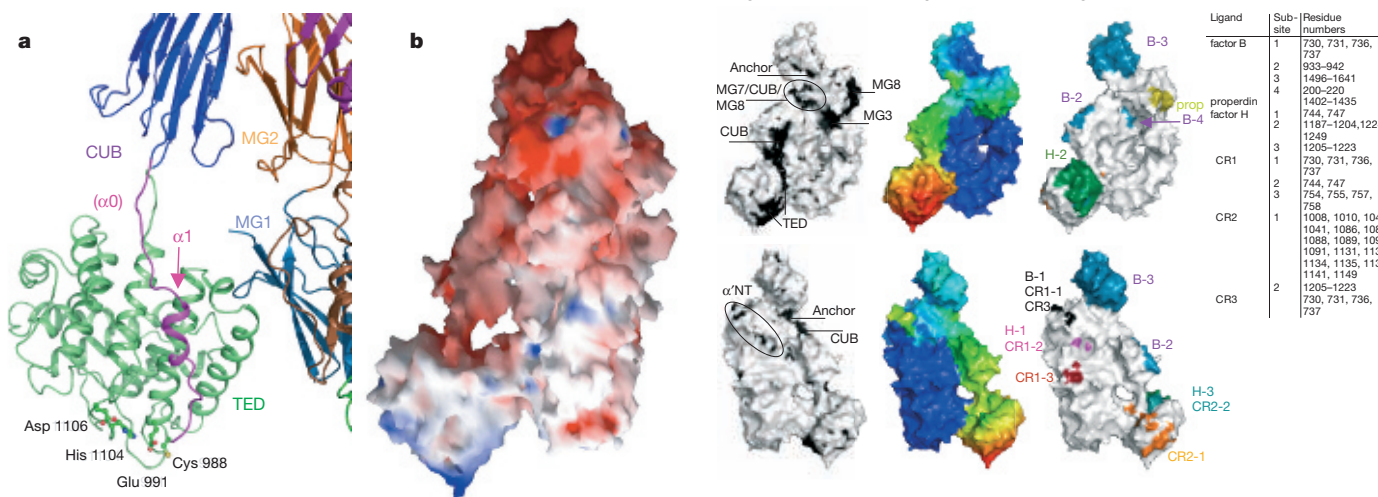


Figure 2 | Thioester exposure for covalent target attachment. **a**, Ribbon representation of the TED domain with the exposed thioester moiety; residues of the acyl-intermediate are shown in ball-and-stick. Indicated in magenta are the regions corresponding to $\alpha 0$ and $\alpha 1$ in C3, which are displaced from the TED domain in C3b. **b**, Surface charge representation³⁰ (shown between -12 (red) and $+12$ kT (blue)) of C3b indicating the overall negative surface-charge potential and the positive patch around the thioester moiety on the TED domain.

Figure 3 | Exposure and generation of cryptic binding sites. **a**, Surface representation of C3b indicating regions exposed in C3b but hidden in C3 (black). **b**, Novel surface arrangements indicated by conformational change (from C3 to C3b) mapped onto the molecular surface, rainbow colouring from minimum (0 Å; blue) to maximum (95 Å; red) coordinate change. **c**, Mapping of the putative binding sites for factor B, properdin, factor H, CR1, CR2 and CR3 as indicated (references as given in the text). Lower panels are rotated 130° relative to upper panels; surface calculated by rolling a sphere of 4-Å radius.

dissociation (regulators of complement). We analysed the C3b surface for exposure of previously hidden interfaces and for new arrangements of sites already exposed on C3 (Fig. 3a, b). The new N-terminal segment α' NT (residues 727–745) carries four acidic residues that are important for binding factor B. Previously we have shown that this segment undergoes a drastic relocation between the native C3 and the final proteolytic fragment C3c (ref. 5). The structure of C3b reveals that this relocation, in which the α' NT moves through the central hole of the β -ring and becomes exposed on the opposite side of the molecule, occurs during the activation of C3 into C3b. Mapping other putative binding sites for factor B onto the C3b surface (Fig. 3c) gives three exposed sub-sites, α' NT¹⁸, strand β 4 of CUB¹⁹ and C345C²⁰ (whereas the putative 'complement C2 receptor inhibitor trispanning' (CRIT) like site²¹ remains mostly occluded), that are located around the shoulder (MG7), neck and head region. This region undergoes significant structural rearrangements from C3 to C3b (Fig. 2b). The absence of novel features on the surface of the β -ring excludes this region as a primary binding site of factor B. In addition, the TED domain is probably not a primary binding site for factor B, because cobra-venom factor, which is a C3b homologue that can make potent fluid-phase convertases, lacks a TED domain.

The association of C3b and factor B is accelerated by properdin, which also stabilizes the resulting convertase C3bBb. The putative binding site of properdin has been mapped to residues 1402–1435 (ref. 22), which form the structurally variable $\beta\alpha\beta$ – $\beta\alpha\alpha$ motif in the MG8 domain. This site is hidden in C3 and becomes exposed in C3b (Fig. 3a). Thus, the arrangement of putative binding sites identifies the arm–shoulder–neck–head region as being responsible for binding factor B and properdin. Regulators of complement (including soluble factor H and cell-surface CR1 and decay-accelerating factor (DAF/CD55)) dissociate the C3bBb complex and act as co-factors in the proteolysis of C3b into iC3b (for example, factor H, CR1 and cell-surface membrane-cofactor protein (MCP/CD46)). Three regions have been implicated in regulator binding: the α' NT (for factor H and CR1), the MG6 domain (for factor H and CR1)²³ and the TED domain (for factor H)²⁴. In the structure of C3b, α' NT and MG6 form one continuous exposed region, indicating that the binding site for factor B overlaps with that for factor H and CR1. Therefore, dissociation of C3bBb by these regulators is achieved through steric hindrance. Binding of factor H to the TED domain might be linked to its cofactor activity. Protease factor I, which degrades C3b by three sequential cleavages (see Supplementary Fig. 2), requires these cofactors (for example, factor H, CR1 or MCP), putatively for orienting its protease domain and unraveling the CUB domain. The resulting iC3b no longer binds factor B. As the arrangement of the 10 domains in C3c is similar to that in C3b, it is likely that iC3b has the same overall core-domain arrangement. Therefore, it seems that the stability of the convertase depends on the structure and orientation of the CUB domain with respect to the rest of C3b.

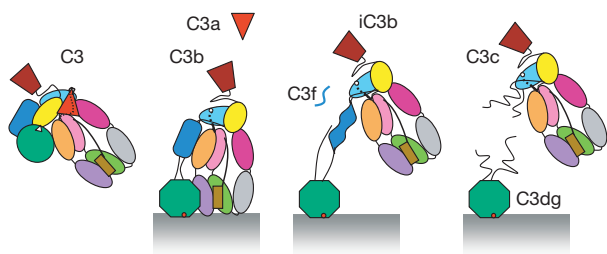


Figure 4 | Proposed model for the conformational pathway of C3. Shown schematically are the four stages: C3, C3b with C3a, iC3b with C3f and C3dg with C3c (see also Supplementary Fig. 2). These conformational changes determine the binding affinities towards soluble proteins (for example, factor B, properdin and factor H) and cell-surface receptors (for example, CR1–4, CRiG, DAF and MCP) that underlie the biological activity.

C3b, iC3b and C3dg induce various cellular responses. C3b and iC3b act as opsonins to facilitate the uptake of tagged particles. This occurs through specific recognition of C3b and iC3b by complement receptors on various phagocytic cells². C3b is recognized by CR1 and CRiG⁹ whereas iC3b is also recognized by the integrins CR3 and CR4. Furthermore, pathogen-bound iC3b and C3dg, but not C3b, contribute to an increased antibody response by binding to CR2 on B cells. A putative binding site for CR2 has been mapped partly on the interface between the TED and MG1 domains²⁵, close to the TED–CUB connections. Therefore it seems likely that, when the TED–CUB connections in iC3b are disrupted, the TED–MG1 interface is opened to provide a new interface for binding CR2 on B cells. Dislodging the iC3b core from the TED domain and the target surface might provide the new signal for binding the integrins CR3 and CR4 on phagocytic cells.

In summary, these data indicate a conformational pathway for C3 and its fragments (Fig. 4) that determines their biological activity. In this pathway, the β -ring provides a stable platform for the functionally important domains of the α -chain that undergo conformational rearrangements. We expect that similar rearrangements occur in homologous host-defence proteins (see Supplementary Fig. 3, where we modelled native α 2-macroglobulin and activated α 2-macroglobulin in electron-microscopy reconstruction maps). Insights into these marked conformational changes are essential for understanding convertase activity and regulation, and will help in the development of complement inhibitors for therapeutic use.

METHODS

Protein purification. C3 was isolated from citrated human plasma as previously described²⁶ with the exception that after the diethylaminoethyl (DEAE) column it was converted to C3b by limited digestion with trypsin for 2 min at 37 °C with 1% w/w enzyme/substrate (the reaction was stopped by the addition of 3% w/w soybean trypsin inhibitor). The C3b was purified immediately by gel filtration over a Superdex 200 size-exclusion column in phosphate buffered saline pH 7.4, followed by anion-exchange chromatography over a Mono-Q column (equilibrated in 50 mM phosphate buffer, pH 7.5 and eluted with 20 column volumes of 0–500 mM NaCl gradient). The protein was then treated with 100 mM iodoacetamide (30 min at 4 °C) to prevent dimerization of the C3b by the free cysteine, concentrated and further purified over a Mono-S column equilibrated in 50 mM phosphate buffer pH 6.0 and eluted with 20 column volumes of 0–500 mM NaCl gradient. The C3b was concentrated to 10 mg ml^{−1}, dialysed against 10 mM Tris pH 7.4 and frozen at −70 °C until crystallization.

Crystallization and crystallography. C3b was crystallized in sitting drops from mother liquor containing 9% w/v PEG-monomethylether 2000, 100 mM sodium acetate, 10 mM taurine and 50 mM Bis-Tris propane pH 7.8 at 30 °C. Crystals grew to 100 × 100 × 60 μ m within 2 weeks. For cryo-protection, the well solution was replaced by mother liquor with increased (to 40% w/v) PEG-monomethylether 2000 and equilibrated for 24 h at 30 °C; and crystals were flash-cooled in liquid nitrogen. Crystals displayed space group $P2_12_12_1$ ($a = 123.9$, $b = 128.5$, $c = 147.0$ Å), contained one molecule per asymmetric unit and diffracted to 4-Å resolution at ESRF beamline ID14-EH4. Diffraction data was processed using MOSFLM/CCP4 (ref. 27) (data statistics are presented in Supplementary Table 1).

Structure determination. C3b was solved by molecular replacement with Phaser²⁸. First, the MG1–MG6 and LNK domains of C3c (pdb code 2A74)⁵ were placed. Second, we placed C3d (pdb code 1C3D)¹³, yielding a large log-likelihood gain (see Supplementary Table 1B) as compared with a loss of log-likelihood when testing the TED domain from C3 (pdb code 2A73)⁵. Finally, we added step-by-step MG8, MG7, α' NT, anchor and C345C from C3c and CUB from C3, respectively, using molecular graphics in Coot²⁹ with rigid-body refinement in Phaser. This yielded a 97% complete model that was completed by model building of the CUB–TED connections (17 residues) and slight adjustments in the other domain–domain connections. Owing to limited resolution only one round of restrained refinement followed by one round of 'translation, liberation and screw-rotation' (TLS) refinement (with individual B-factors set to 50 Å²) were performed in REFMAC²⁷. The final refined model of C3b, consisting of 1,531 residues, had R and R_{free} values of 27.3 and 32.3%, respectively (for model statistics see Supplementary Table 1). Electron density for each domain is shown in Supplementary Fig. 1; see also Fig. 1b. The central domains fitted excellently in the electron density; the outermost domains (CUB, TED and C345C) showed

convincing, but weaker, electron density. All molecular graphics figures, except Fig. 2b, were generated with pymol (W. Delano; <http://www.pymol.org/>).

Received 9 July; accepted 15 August 2006.

Published online 15 October 2006.

- Walport, M. J. Complement. First of two parts. *N. Engl. J. Med.* **344**, 1058–1066 (2001).
- Carroll, M. C. The complement system in regulation of adaptive immunity. *Nature Immunol.* **5**, 981–986 (2004).
- Law, S. K. & Levine, R. P. Interaction between the third complement protein and cell surface macromolecules. *Proc. Natl Acad. Sci. USA* **74**, 2701–2705 (1977).
- Lambris, J. D. The multifunctional role of C3, the third component of complement. *Immunol. Today* **9**, 387–393 (1988).
- Janssen, B. J. *et al.* Structures of complement component C3 provide insights into the function and evolution of immunity. *Nature* **437**, 505–511 (2005).
- Mastellos, D. & Lambris, J. D. Complement: more than a 'guard' against invading pathogens? *Trends Immunol.* **23**, 485–491 (2002).
- Tack, B. F., Harrison, R. A., Janatova, J., Thomas, M. L. & Pahl, J. W. Evidence for presence of an internal thioester bond in third component of human complement. *Proc. Natl Acad. Sci. USA* **77**, 5764–5768 (1980).
- Fishelson, Z., Pangburn, M. K. & Muller-Eberhard, H. J. Characterization of the initial C3 convertase of the alternative pathway of human complement. *J. Immunol.* **132**, 1430–1434 (1984).
- Helmy, K. Y. *et al.* CRlg: a macrophage complement receptor required for phagocytosis of circulating pathogens. *Cell* **124**, 915–927 (2006).
- Dempsey, P. W., Allison, M. E., Akkaraju, S., Goodnow, C. C. & Fearon, D. T. C3d of complement as a molecular adjuvant: bridging innate and acquired immunity. *Science* **271**, 348–350 (1996).
- Ross, G. D., Lambris, J. D., Cain, J. A. & Newman, S. L. Generation of three different fragments of bound C3 with purified factor I or serum. I. Requirements for factor H vs CR1 cofactor activity. *J. Immunol.* **129**, 2051–2060 (1982).
- Harrison, R. A. & Lachmann, P. J. Novel cleavage products of the third component of human complement. *Mol. Immunol.* **17**, 219–228 (1980).
- Nagar, B., Jones, R. G., Diefenbach, R. J., Isenman, D. E. & Rini, J. M. X-ray crystal structure of C3d: a C3 fragment and ligand for complement receptor 2. *Science* **280**, 1277–1281 (1998).
- Brown, K. M. *et al.* Influence of donor C3 allotype on late renal-transplantation outcome. *N. Engl. J. Med.* **354**, 2014–2023 (2006).
- Dodds, A. W., Ren, X. D., Willis, A. C. & Law, S. K. The reaction mechanism of the internal thioester in the human complement component C4. *Nature* **379**, 177–179 (1996).
- Pangburn, M. K., Schreiber, R. D. & Muller-Eberhard, H. J. C3b deposition during activation of the alternative complement pathway and the effect of deposition on the activating surface. *J. Immunol.* **131**, 1930–1935 (1983).
- Sim, R. B., Twose, T. M., Paterson, D. S. & Sim, E. The covalent-binding reaction of complement component C3. *Biochem. J.* **193**, 115–127 (1981).
- Taniguchi-Sidle, A. & Isenman, D. E. Interactions of human complement component C3 with factor B and with complement receptors type 1 (CR1, CD35) and type 3 (CR3, CD11b/CD18) involve an acidic sequence at the N-terminus of C3 α '-chain. *J. Immunol.* **153**, 5285–5302 (1994).
- O'Keefe, M. C., Caporale, L. H. & Vogel, C. W. A novel cleavage product of human complement component C3 with structural and functional properties of cobra venom factor. *J. Biol. Chem.* **263**, 12690–12697 (1988).
- Kolln, J., Spillner, E., Andra, J., Klensang, K. & Bredehorst, R. Complement inactivation by recombinant human C3 derivatives. *J. Immunol.* **173**, 5540–5545 (2004).
- Inal, J. M. & Schifferli, J. A. Complement C2 receptor inhibitor trispanning and the beta-chain of C4 share a binding site for complement C2. *J. Immunol.* **168**, 5213–5221 (2002).
- Daoudaki, M. E., Becherer, J. D. & Lambris, J. D. A 34-amino acid peptide of the third component of complement mediates properdin binding. *J. Immunol.* **140**, 1577–1580 (1988).
- Oran, A. E. & Isenman, D. E. Identification of residues within the 727–767 segment of human complement component C3 important for its interaction with factor H and with complement receptor 1 (CR1, CD35). *J. Biol. Chem.* **274**, 5120–5130 (1999).
- Lambris, J. D., Avila, D., Becherer, J. D. & Muller-Eberhard, H. J. A discontinuous factor H binding site in the third component of complement as delineated by synthetic peptides. *J. Biol. Chem.* **263**, 12147–12150 (1988).
- Clemenza, L. & Isenman, D. E. Structure-guided identification of C3d residues essential for its binding to complement receptor 2 (CD21). *J. Immunol.* **165**, 3839–3848 (2000).
- Lambris, J. D., Dobson, N. J. & Ross, G. D. Release of endogenous C3b inactivator from lymphocytes in response to triggering membrane receptors for beta 1H globulin. *J. Exp. Med.* **152**, 1625–1644 (1980).
- The C. C. P. 4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763 (1994).
- Storoni, L. C., McCoy, A. J. & Read, R. J. Likelihood-enhanced fast rotation functions. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 432–438 (2004).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
- Nicholls, A., Bharadwaj, R. & Honig, B. GRASP: Graphical representation and analysis of surface properties. *Biophys. J.* **64**, 166–170 (1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge the European Synchrotron Radiation Facility for provision of synchrotron radiation facilities; and S.J. Kolodziej and J.K. Stoops (Houston) for providing the native α 2-macroglobulin and methylamine-treated α 2-macroglobulin EM-reconstruction maps. We thank B. Nilsson (Uppsala), R.B.G. Ravelli (Grenoble) and E.G. Huizinga (Utrecht) for reading the manuscript. This work was supported by a 'Pioneer' programme grant to P.G. by the Council for Chemical Sciences of the Netherlands Organization for Scientific Research (NWO-CW) and an NIH grant to J.D.L.

Author Information Coordinates and structure factors of the C3b structure have been deposited in the Protein Data Bank with accession number 2I07. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.G. (p.gros@chem.uu.nl).

Structure of C3b in complex with CRiG gives insights into regulation of complement activation

Christian Wiesmann¹, Kenneth J. Katschke Jr², JianPing Yin¹, Karim Y. Helmy², Micah Steffek³, Wayne J. Fairbrother¹, Scott A. McCallum¹, Lizette Embuscado⁴, Laura DeForge⁴, Philip E. Hass³ & Menno van Lookeren Campagne²

The complement system is a key part of the innate immune system, and is required for clearance of pathogens from the bloodstream¹. After exposure to pathogens, the third component of the complement system, C3, is cleaved to C3b which, after recruitment of factor B, initiates formation of the alternative pathway convertases^{2–4}. CRiG, a complement receptor expressed on macrophages, binds to C3b and iC3b mediating phagocytosis of the particles⁵, but it is unknown how CRiG selectively recognizes proteolytic C3-fragments and whether binding of CRiG to C3b inhibits convertase activation. Here we present the crystal structure of C3b in complex with CRiG and, using CRiG mutants, provide evidence that CRiG acts as an inhibitor of the alternative pathway of complement. The structure shows that activation of C3 induces major structural rearrangements, including a dramatic movement (>80 Å) of the thioester-bond-containing domain through which C3b attaches to pathogen surfaces. We show that CRiG is not only a phagocytic receptor, but also a potent inhibitor of the alternative pathway convertases. The structure provides insights into the complex macromolecular structural rearrangements that occur during complement activation and inhibition. Moreover, our structure–function studies relating the structural basis of complement activation and the means by which CRiG inhibits the convertases provide important clues to the development of therapeutics that target complement.

Component C3 has a key role in complement activation through the classical, lectin and alternative pathways of complement¹. C3 is activated following cleavage by the proteolytic enzyme 'C3 convertase'^{6,7}, generating the small (9 kDa) C3a peptide, also called anaphylatoxin, and the large (177 kDa) C3b fragment, which covalently attaches to particles before degradation to iC3b and C3c. CRiG recognizes these molecules but not parent C3. To understand this specificity of CRiG, we determined the crystal structure of human CRiG in its unbound state, in complex with C3b as well as C3c, at 1.2-, 4.1- and 3.1-Å resolution, respectively. In terms of size, C3b is more similar to the parent C3 than to the breakdown fragment C3c. On a structural level, however, the core of C3b is identical to C3c. Briefly, like C3, C3b is composed of two chains. The β -chain (residues 1–645) folds into five macroglobulin-like domains (MG1–MG5), the amino-terminal half of a sixth MG domain, and a linker region (LNK) (Fig. 1). These MG domains are arranged in a 'key-ring'-like fashion and circle around a central cavity with dimensions of 10 Å × 30 Å. The N-terminal residues of the C3b α -chain (727–1641) are referred to as the α' NT domain (727–745) and adopt an extended conformation. Next, residues 746–807, together with residues 534–578 from the β -chain, complete the sixth MG domain. Two additional MG domains (MG7 and MG8) are separated by the CUB domain (protein

module initially found in complement subcomponents C1r/C1s, Uegf and Bone morphogenetic protein-1) and the thioester-bond-containing domain (TED) (Fig. 1), and the carboxy-terminal 170 residues form the 'anchor region'⁸ and the C345C domain. The core of C3b and C3c are virtually identical; the 642 C α s of the β -chain and the 479 C α s of the α -chain superimpose with root mean square deviations (RMSDs) of 0.6 Å and 0.5 Å, respectively. Thus, the conformational changes between C3 and C3c noted in ref. 8 also apply to the comparison between C3 and C3b.

In contrast to the structural similarities between C3b and C3c, comparison between C3b and C3 reveals that C3 activation is accompanied by major rearrangements (Fig. 1, Supplementary Fig. 1). During C3 cleavage, the release of the anaphylatoxin (ANA) domain, which in C3 is wedged between MG3 and MG8 (ref. 9), triggers a rotation of the latter two domains, resulting in a rearrangement of MG7 and the intriguing movement of the α' NT domain⁸. In C3, the α' NT domain is mostly buried. In C3b, this domain has become solvent-accessible, exposing a number of potential binding sites for receptors and regulators of complement activation, such as complement receptor 1, factor H, and factor B (refs 10–12). Even more dramatically, the CUB and TED domains have undergone major and unanticipated movements in C3b (Supplementary Fig. 1). These domain movements serve to expose and position the thioester 'warhead'¹³ responsible for the covalent attachment of C3b to pathogens. In C3, the thioester group is thoroughly shielded from solvent and embedded in a hydrophobic pocket formed by domains from the α -chain⁸. Furthermore, local restraints imposed by the interactions of the thioester domain with the CUB and MG8 domains prevent the formation of the critical acyl-imidazol intermediate resulting from the attack of Gln 991 by His 1104 in C3b (ref. 14). In C3b, the 'warhead' is found at a distance of about 80 Å from its original position, ending up close to the β -chain, and is completely solvent-exposed, protruding from the surface. The amino acids important for covalent attachment (Cys 988, Gln 991, His 1104) are now adjacent to each other and positioned to attack particle surfaces or to be hydrolysed by water.

As predicted¹⁵, the N-terminal domain of CRiG belongs to the IgV family of immunoglobulin-like domains. Like all members of this family, the CRiG domain contains two β -sheets, one composed of strands A', G, F, C, C' and C'', and the other of strands B, E and D (Fig. 2a, b, and Supplementary Fig. 3a). CRiG engages C3b and C3c in an identical fashion (Fig. 1). In line with this, the affinities of CRiG for C3b and C3c are similar (Supplementary Fig. 2). Comparison of the C3c and CRiG structures in their unbound states with the complex reveals that binding does not involve significant conformational changes. The largest difference occurs within flexible regions, with

¹Department of Protein Engineering, ²Department of Immunology, ³Department of Protein Chemistry, ⁴Department of Assay Technology, Genentech Inc., 1 DNA Way South San Francisco, California 94080, USA.

only one small loop movement in the MG6 domain. The CRiG binding site and the TED domain lie on opposite sites of the β -chain, separated by 40 Å (Fig. 1). Thus, macrophage-expressed CRiG can easily bind to C3b that is covalently attached to pathogens.

The binding interface between CRiG and C3c or C3b is large, discontinuous, and buries 2,670 Å² of solvent-accessible surface (Fig. 2a, b). CRiG's binding site is composed of residues from strands A', B, C', C'', F and G, with the majority of interactions formed by residues on or near strands C' and C'' (Supplementary Fig. 3a). The hairpin loop connecting C' and C'' protrudes into the cavity in the centre of the key-ring-shaped β -chain of C3b (Fig. 2a). Unlike any other complement receptors identified to date^{10–12,16}, CRiG binds primarily to the β -chain of C3b. Domains MG3, MG4, MG5, MG6 and LNK all interact with CRiG, with the largest contributions being made by MG3 and MG6 at 30% and 40%, respectively, of the total buried interface (Fig. 2a, b, Supplementary Fig. 3b). MG3 has rotated by 15° when comparing C3 to C3c (ref. 8) and C3b (Supplementary Fig. 4). The C3b:CRiG structure shows that this reorientation of MG3, along with the movement of a helical section in the LNK region, is necessary to form the CRiG binding site, thus explaining why CRiG cannot bind to C3 (ref. 5).

As C3b forms a central subunit of the complement convertases, we investigated if CRiG binding to C3b influences C3 and C5 convertase activity. C3 together with factors B and D generates the convertases of the alternative pathway (AP)^{4,17}. In the presence of CRiG, but not of a

control protein, generation of C3a and C3b by C3 convertase and generation of C5b by C5 convertase were inhibited (Fig. 2c, d and Supplementary Fig. 5a).

To confirm the relevance of the binding interface between CRiG and C3b revealed by the crystal structure, and to determine the effect of CRiG binding on convertase activity, we generated several CRiG mutants. Mutation of residues that are in contact with MG3, MG5 or MG6 (mutation A, B, and C; Fig. 3a, Supplementary Fig. 3a) in the crystal structure effectively abrogated binding of CRiG to C3b and inhibition of AP haemolytic activity (half-maximal inhibitory concentration (IC₅₀) IC₅₀ values were increased >50 fold as compared to wild-type CRiG (Fig. 3b and c)). Mutation of residues peripheral to the interface (mutants D and E) modestly decreased binding affinity and inhibition of haemolytic activity (10–50 fold increase in IC₅₀ values), whereas changing residues outside the binding interface (control mutant) did not affect C3b binding affinity or inhibition of haemolytic activity. The potency of these mutants to inhibit the AP in the haemolytic assay (Supplementary Fig. 5b). In conclusion, CRiG inhibition of the AP depends on its ability to bind C3b.

The C3b subunit of the convertases does not have protease activity. Rather, C3b is responsible for the recruitment of C3 and C5, thus enabling the catalytic subunit factor Bb to cleave these substrates^{18,19}. Unlike many other complement regulators, CRiG does not inhibit convertase activity by displacing factor Bb from the C3b molecule (decay activity; Supplementary Fig. 6a), nor does it act as a co-factor for factor I mediated proteolysis of C3b (co-factor activity; Supplementary Fig. 6b). Therefore, CRiG must inhibit the convertase through another mechanism, such as interference with the substrate binding function of C3b. To address this, we tested the ability of

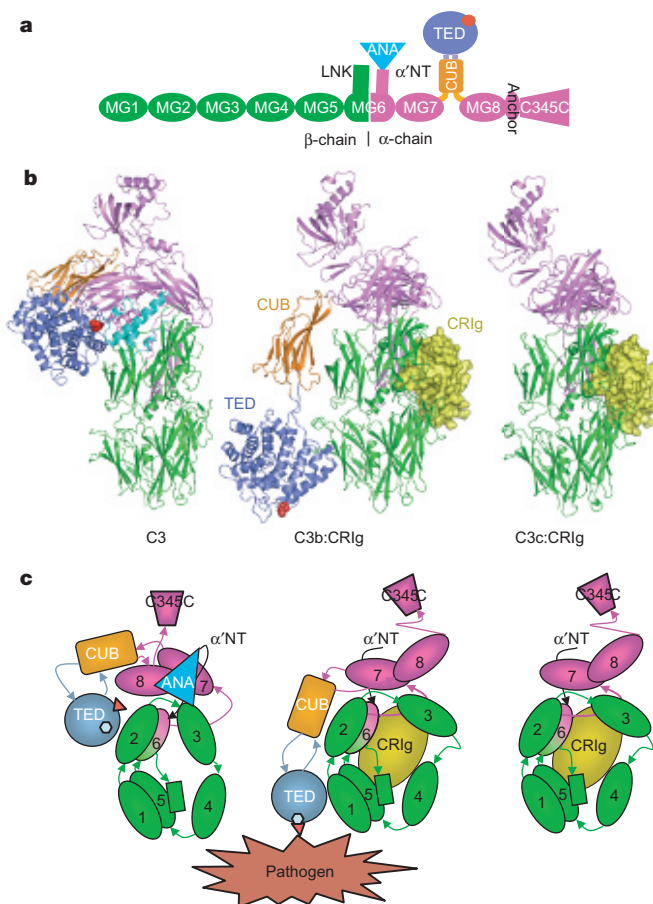


Figure 1 | Domain architecture and structure of C3 and the C3b/C3c:CRiG complexes. **a**, Domain organization of C3; the β -chain is depicted in green, the α -chain in cyan (ANA domain), orange, (CUB), blue (TED) and violet. The red sphere indicates the thioester. **b**, Ribbon diagrams of native C3; (left; ref. 8), C3b in complex with CRiG (centre) and C3c in complex with CRiG (right). The surface of CRiG is shown in yellow. Note the movement of the CUB and TED domain when comparing C3 and C3b:CRiG. **c**, Schematic depiction of **b** from a slightly different orientation; the hexagon and triangle represent His 1104 and the residues that form the thioester bond in C3, respectively.

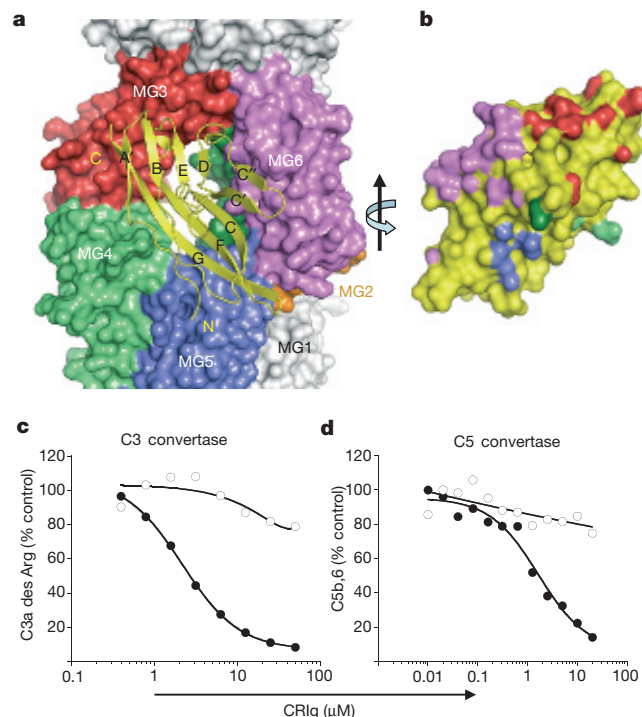


Figure 2 | CRiG binds to the β -chain of C3b and C3c and inhibits the AP C3 and C5 convertases. **a**, Complex between C3c (surface representation) and CRiG (ribbon diagram). **b**, Surface representation of CRiG, rotated 180° from **a**. All atoms that are close to C3c/C3b are coloured like the residues they are contacting in **a**. **c**, CRiG (filled symbols) but not control protein (open symbols) reduces formation of C3a des Arg formed on cleavage of C3 to C3a and C3b. **d**, CRiG (filled symbols) but not control protein (open symbols) inhibits generation of C5b, 6 formed on cleavage of C5 to C5a and C5b. Data are representative of three independent experiments.

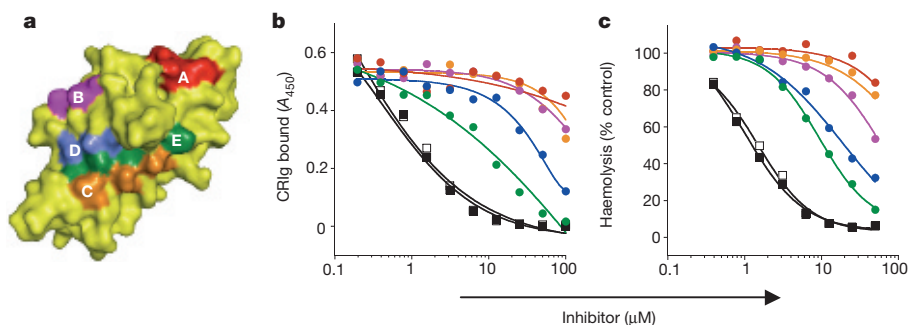


Figure 3 | Structure–activity relationship of CRiG mutants. **a**, Surface of CRiG in the same orientation as in Fig. 2b. CRiG mutants used for **b** and **c** are coloured. Mutants are described in Supplementary Fig. 3a. **b**, The relative affinity of CRiG mutants was determined by competition with CRiG fused to

(C3b)₂ to bind C5 in the presence of CRiG or control proteins. Here, only wild-type CRiG blocks C5 binding to (C3b)₂ in a concentration dependent manner (Fig. 4a, left panel). Furthermore, CRiG inhibits AP C5 convertase activity, as increased amounts of CRiG prevent the formation of C5b-9 in an assay selective for the AP²⁰ (Fig. 4a, right panel). We next determined if CRiG was able to compete with C5 binding to the heterodimeric C3bC4b subunit of the classical pathway (CP) convertase. Although CRiG bound to the heterodimeric subunit with an affinity similar to that of a monomeric C3b molecule (results not shown), CRiG did not inhibit C5 binding to the CP C5 convertase, as C5 binding to the C3bC4b heterodimer was unaffected by the presence of CRiG (Fig. 4b, left panel). In line with the absence of any effect on C5 binding to a C3bC4b heterodimer, CRiG did not alter the formation of C5b-9 in an assay selective

for the CP (Fig. 4b, right panel). In addition, haemolytic assays selective for the CP failed to demonstrate inhibition by CRiG (Supplementary Fig. 7). These results show that binding of CRiG to the β -chain of C3b selectively abrogates the interaction of C3 and C5 with the AP but not with the CP convertases.

In conclusion, CRiG, a novel Ig-superfamily complement receptor, binds primarily to the β -chain of C3b and C3c in a location that is distinct from the putative binding sites of other complement regulators. CRiG binding to C3b results in inhibition of complement activation through the AP. Our structure–function studies provide important clues for the development of therapeutics that selectively target the alternative complement pathway.

for the CP (Fig. 4b, right panel). In addition, haemolytic assays selective for the CP failed to demonstrate inhibition by CRiG (Supplementary Fig. 7). These results show that binding of CRiG to the β -chain of C3b selectively abrogates the interaction of C3 and C5 with the AP but not with the CP convertases.

METHODS

All complement proteins were obtained from CompTech except C3 and its fragments, which were generated as described in Supplementary Information. Protein expression, purification, crystallization, data collection and refinement are described in Supplementary Information. The effect of CRiG on fluid-phase C3 cleavage by C3 convertase was performed according to ref. 21 with modifications as detailed in Supplementary Methods. The C5 convertase assay was performed as described elsewhere²². Haemolysis through the AP and CP of complement were assessed with rabbit erythrocytes (Er, Colorado Serum) or IgM-coated sheep erythrocytes (E-IgM, CompTech) as described^{21,23} with modifications as detailed in Supplementary Methods. Microtitre assays for the CP and AP were performed with the Wieslab Complement system screen (Alpco Diagnostics) according to the instructions supplied with the kit. Factor B- and C2-depleted human serum was used to measure activity of the CP and AP, respectively. Relative affinities of the various CRiG proteins to C3b were determined by incubating C3b-coated microtitre plates with increasing concentrations of CRiG proteins and a fixed concentration (400 pM) of CRiG-LFH. CRiG-LFH binding was detected with anti-Flag M2 antibody conjugated to HRPO (Sigma). Plates were developed with TMB substrate solution (KPL) and the reaction stopped with 2 N H₂SO₄. Absorbance was measured at 450 nm. Two wells were averaged and considered a single data point. IC₅₀ values were determined by nonlinear regression analysis using a four parameter fit formula (Synergy Software).

Received 21 August; accepted 21 September 2006.
Published online 15 October 2006.

- Walport, M. J. Complement. First of two parts. *N. Engl. J. Med.* **344**, 1058–1066 (2001).
- Pangburn, M. K. & Muller-Eberhard, H. J. Relation of putative thioester bond in C3 to activation of the alternative pathway and the binding of C3b to biological targets of complement. *J. Exp. Med.* **152**, 1102–1114 (1980).
- Fishelson, Z., Pangburn, M. K. & Muller-Eberhard, H. J. C3 convertase of the alternative complement pathway. Demonstration of an active, stable C3b, Bb (Ni) complex. *J. Biol. Chem.* **258**, 7411–7415 (1983).
- Muller-Eberhard, H. J. & Schreiber, R. D. Molecular biology and chemistry of the alternative pathway of complement. *Adv. Immunol.* **29**, 1–53 (1980).
- Helmy, K. Y. *et al.* CRiG: a macrophage complement receptor required for phagocytosis of circulating pathogens. *Cell* **124**, 915–927 (2006).
- Pangburn, M. K., Schreiber, R. D. & Muller-Eberhard, H. J. Formation of the initial C3 convertase of the alternative complement pathway. Acquisition of C3b-like activities by spontaneous hydrolysis of the putative thioester in native C3. *J. Exp. Med.* **154**, 856–867 (1981).

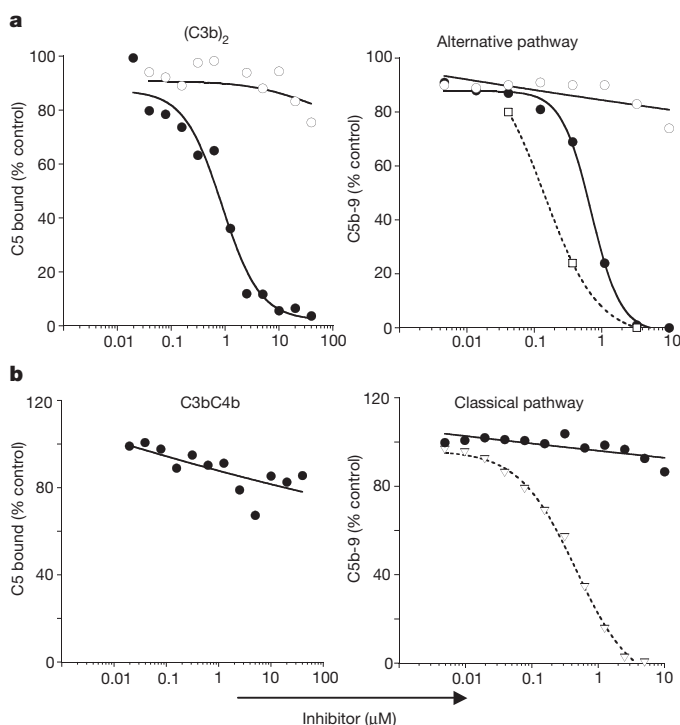


Figure 4 | CRiG inhibits C5 binding to the AP C5 convertase. **a**, Left panel: CRiG blocks C5 binding to (C3b)₂. (C3b)₂, captured on a microtitre plate, was incubated with C5 and increasing concentrations of CRiG (filled symbols) or CRiG mutant A (open symbols). Right panel: CRiG inhibits C5b-9 generation in an enzyme immunoassay selective for the AP. **b**, Left panel: CRiG (filled symbols) does not inhibit C5 binding to C3bC4b captured on a microtitre plate. Right panel: CRiG does not inhibit CP convertase activation in an enzyme immunoassay selective for the CP. The dashed line in **a** and **b** represents factor H (**a**) and C1 inhibitor (C1NH; **b**). Data are representative of three independent experiments.

7. Sahu, A. & Lambris, J. D. Structure and biology of complement protein C3, a connecting link between innate and acquired immunity. *Immunol. Rev.* **180**, 35–48 (2001).
 8. Janssen, B. J. *et al.* Structures of complement component C3 provide insights into the function and evolution of immunity. *Nature* **437**, 505–511 (2005).
 9. Fredslund, F. *et al.* The structure of bovine complement component 3 reveals the basis for thioester function. *J. Mol. Biol.* **361**, 115–127 (2006).
 10. Lambris, J. D. *et al.* Dissection of CR1, factor H, membrane cofactor protein, and factor B binding and functional sites in the third complement component. *J. Immunol.* **156**, 4821–4832 (1996).
 11. Taniguchi-Sidle, A. & Isenman, D. E. Interactions of human complement component C3 with factor B and with complement receptors type 1 (CR1, CD35) and type 3 (CR3, CD11b/CD18) involve an acidic sequence at the N-terminus of C3 α' -chain. *J. Immunol.* **153**, 5285–5302 (1994).
 12. Oran, A. E. & Isenman, D. E. Identification of residues within the 727–767 segment of human complement component C3 important for its interaction with factor H and with complement receptor 1 (CR1, CD35). *J. Biol. Chem.* **274**, 5120–5130 (1999).
 13. Liddington, R. & Bankston, L. Structural biology: origins of chemical biodefence. *Nature* **437**, 484–485 (2005).
 14. Gadjeva, M. *et al.* The covalent binding reaction of complement component C3. *J. Immunol.* **161**, 985–990 (1998).
 15. Langnaese, K., Colleaux, L., Kloos, D. U., Fontes, M. & Wieacker, P. Cloning of Z39Ig, a novel gene with immunoglobulin-like domains located on human chromosome X. *Biochim. Biophys. Acta* **1492**, 522–525 (2000).
 16. Janssen, B. J. & Gros, P. Structural insights into the central complement component C3. *Mol. Immunol.* (in the press).
 17. Pangburn, M. K. & Muller-Eberhard, H. J. The C3 convertase of the alternative pathway of human complement. Enzymic properties of the bimolecular proteinase. *Biochem. J.* **235**, 723–730 (1986).
 18. Vogt, W., Schmidt, G., Von Buttlar, B. & Dieminger, L. A new function of the activated third component of complement: binding to C5, an essential step for C5 activation. *Immunology* **34**, 29–40 (1978).
 19. Rawal, N. & Pangburn, M. K. Functional role of the noncatalytic subunit of complement C5 convertase. *J. Immunol.* **164**, 1379–1385 (2000).
 20. Seelen, M. A. *et al.* Functional analysis of the classical, alternative, and MBL pathways of the complement system: standardization and validation of a simple ELISA. *J. Immunol. Methods* **296**, 187–198 (2005).
 21. Kostavasili, I. *et al.* Mechanism of complement inactivation by glycoprotein C of herpes simplex virus. *J. Immunol.* **158**, 1763–1771 (1997).
 22. Rawal, N. & Pangburn, M. K. C5 convertase of the alternative pathway of complement. Kinetic analysis of the free and surface-bound forms of the enzyme. *J. Biol. Chem.* **273**, 16828–16835 (1998).
 23. Krych-Goldberg, M. *et al.* Decay accelerating activity of complement receptor type 1 (CD35). Two active sites are required for dissociating C5 convertases. *J. Biol. Chem.* **274**, 31160–31168 (1999).
- Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.
- Acknowledgements** We thank B. de Vos, A. Chan, F. Bazan and H. Spits for critically reading the manuscript, and J. Chinn and B. Appleton for important contributions. We acknowledge the support staff at beamline 5.0.2 at the Advanced Light Source. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences Division of the US Department of Energy at Lawrence Berkeley National Laboratory.
- Author Contributions** K.J.K., J.P.Y., K.Y.H. and M.S. contributed equally to this work. K.J.K., L.E. and L.D.F. performed the C3 convertase and haemolysis assay, S.A.M.C. and W.J.F. contributed to mutant design, K.Y.H. generated CR1g mutants and performed the microwell assays, M.S. and P.E.H. purified the complement proteins, J.P.Y. generated CR1g for crystallography, purified the complexes and grew the crystals, C.W. solved the structures and wrote the paper together with M.v.L.C.
- Author Information** Coordinates and structure factors have been deposited at the RCSB Protein Databank (2ICC, 2ICE, and 2ICF). Reprints and permission information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.W. (wiesmann.christian@gene.com) or M.v.L.C. (menno@gene.com).

The structure of complement C3b provides insights into complement activation and regulation

A. Abdul Ajees¹, K. Gunasekaran¹, John E. Volanakis², Sthanam. V. L. Narayana¹, Girish J. Kotwal³ & H. M. Krishna Murthy¹

The human complement system is an important component of innate immunity. Complement-derived products mediate functions contributing to pathogen killing and elimination¹. However, inappropriate activation of the system contributes to the pathogenesis of immunological and inflammatory diseases¹. Complement component 3 (C3) occupies a central position because of the manifold biological activities of its activation fragments, including the major fragment, C3b, which anchors the assembly of convertases effecting C3 and C5 activation. C3 is converted to C3b by proteolysis of its anaphylatoxin domain², by either of two C3 convertases. This activates a stable thioester bond, leading to the covalent attachment of C3b to cell-surface or protein-surface hydroxyl groups through transesterification³. The cleavage and activation of C3 exposes binding sites for factors B, H and I, properdin, decay accelerating factor (DAF, CD55), membrane cofactor protein (MCP, CD46), complement receptor 1 (CR1, CD35) and viral molecules such as vaccinia virus complement-control protein⁴. C3b associates with these molecules in different configurations and forms complexes mediating the activation, amplification and regulation of the complement response^{1,4}. Structures of C3 and C3c, a fragment derived from the proteolysis of C3b, have revealed a domain configuration, including six macroglobulin domains (MG1–MG6; nomenclature follows ref. 5) arranged in a ring, termed the β -ring⁵. However, because neither C3 nor C3c is active in complement activation and regulation, questions about function can be answered only through direct observations on C3b. Here we present a structure of C3b that reveals a marked loss of secondary structure in the CUB (for ‘complement C1r/C1s, Uegf, Bmp1’) domain, which together with the resulting translocation of the thioester domain provides a molecular basis for conformational changes accompanying the conversion of C3 to C3b. The total conformational changes make many proposed ligand-binding sites more accessible and create a cavity that shields target peptide bonds from access by factor I. A covalently bound *N*-acetyl-L-threonine residue demonstrates the geometry of C3b attachment to surface hydroxyl groups.

Large conformational changes in several domains are pivotal to the conversion of inactive C3 to active C3b. A major contributor to the structural stability of C3 is MG8, which is in close contact with the MG3, anaphylatoxin (ANA), amino-terminal region of the cleaved α -chain (residues 727–745) referred as α' NT (ANT), MG7, CUB and thioester (TED) domains, burying 198, 650, 90, 629, 180 and 1,002 Å² of accessible surface area, respectively (Fig. 1a). Proteolytic removal of ANA thus destabilizes MG8. In addition, MG3 and ANT, which bury 244 and 395 Å² of accessible area at the C3–ANA interface, are also destabilized. Neighbouring β -ring domains limit the movement of MG3 to a rotation of 15° and a translation of 4 Å. However, ANT

rotates through 163° and flips to a new position 37 Å away (Fig. 1a), further destabilizing MG8. As a result, MG8 rotates through 66° and translates by 24 Å, veering away from CUB and TED (Fig. 1a). TED, released from packing confines, rotates through 164° and translates by 84 Å. TED movement is facilitated by the unravelling of the CUB β -sheet. In C3, CUB, an eight-stranded β -sheet, is composed of two half domains, namely CUBg (residues 912–962, strands β 1– β 5) and CUBf (1269–1330, β 6– β 8), constituting, respectively, part of C3g (933–979) and C3f (1281–1299). CUB and TED are iteratively inserted into the rest of C3: CUB between MG7 and MG8, and TED within the β 5– β 6 loop of CUB⁵. The N terminus of TED is thus connected to CUBg and its carboxy terminus to CUBf. In C3b, both CUBg and CUBf lose their β -sheet structure, with consequent spatial separation. CUBg elongates into a one-turn α -helix, a type I β -turn and a succession of 13 type IV β -turns, whereas CUBf extends into five successive type IV β -turns and two tandem γ -turns (Fig. 1b). Stretching of CUB subdomains in C3 is probably aided by lack of the disulphide bonds and metal-binding sites that stabilize CUB domains in other proteins⁶. Cumulatively, these structural changes make C3b more elongated and open than C3 (Fig. 2, and Supplementary Movie). As a direct consequence, putative binding sites for large ligands, such as factor B and properdin, are spatially dispersed, reducing potential steric collisions between them. Other C3 domains remain essentially unaltered in C3b and C3c. The overall folding of C3d is also conserved in C3, C3b, human⁷ and rat⁸ homologues.

C3b is a key component of all three pathways of complement activation. In the alternative pathway (AP), C3b provides the non-catalytic subunits of both C3 convertase and C5 convertases, whereas in the classical (CP) and lectin (LP) pathways it provides one of two noncatalytic subunits of the C5 convertase. Covalent attachment of C3b to an AP-activating surface forms the nidus of the C3 convertase, whereas its attachment to a C3 convertase changes the latter's specificity to C5 convertase. Both attachments are mediated through the reactive carbonyl of Gln 991, which, serving as the acyl donor, esterifies hydroxyl groups on acceptors. In C3, Gln 991 forms a buried thioester with Cys 988; on removal of ANA, His 1104 attacks this thioester to form an acyl-imidazole intermediate, freeing the thiolate of Cys 988, which catalyses the transfer of the acyl group of Gln 991 to an acceptor hydroxyl¹. His 1104 in C3 is too far (11.7 Å) from the thioester to form the acyl-imidazole intermediate, and its approach is blocked by the TED–MG8 interface⁵. In C3b, proteolytic removal of ANA has evidently resulted in a rotation of TED, fully exposing the thioester to solvent and eliminating the hindrance to the approach of His 1104 to Cys 988 and Gln 991. The resulting activated acyl group of Gln 991 has esterified the γ -hydroxyl of the *N*-acetyl-L-threonine (AcT) present in the medium. The attached AcT (Supplementary Fig. 1) is geometrically similar to free AcT (ref. 9), except at the site

¹Center for Biophysical Sciences and Engineering, ²Department of Medicine, University of Alabama at Birmingham, 1530 3rd Avenue South Birmingham, Alabama 35294, USA.

³Division of Medical Virology, Institute for Infectious Diseases and Molecular Medicine, University of Cape Town, HSC, Cape Town 7925, South Africa.

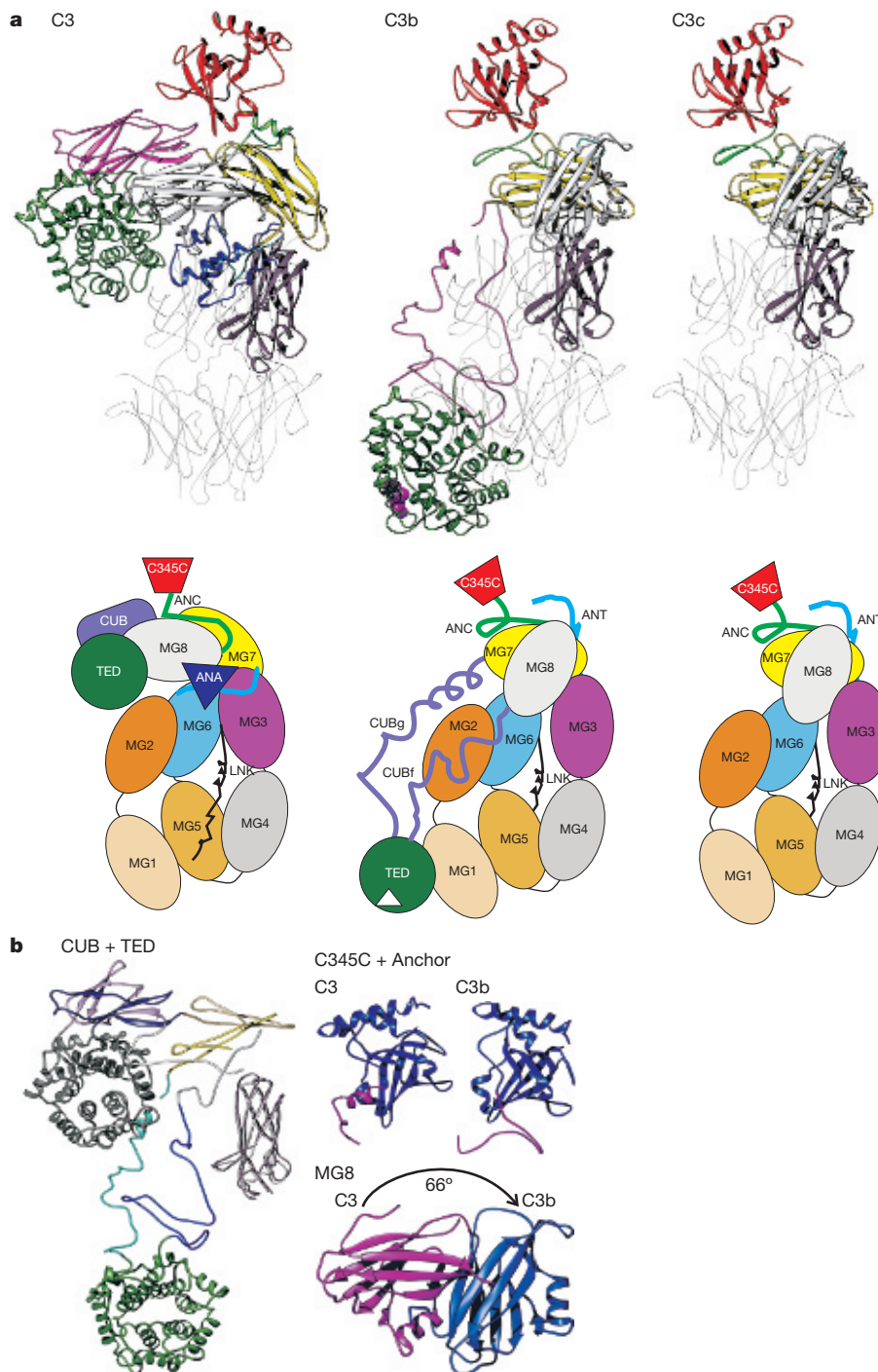


Figure 1 | C3, C3b and C3c. **a**, Top: diagrams showing domain movement on conversion to C3b. Red, C345C; yellow, MG7; grey, MG8; blue, ANA; purple, MG3; violet, CUBg and CUBf; green, TED. Bottom: schematic representation of the diagrams in the top row. The AcT-binding site is indicated by a white arrowhead. LNK, linker domain (residues 578–645).

of attachment to Gln 991. The magnitude of TED translation is comparable to the more than 50 Å estimated from X-ray scattering measurements¹. One energy source driving structural changes between C3 and C3b might be hydrolysis of the high-energy buried thioester bond⁵, in a manner similar to that in α_2 -macroglobulin, a related protein¹⁰. Residues 944–1273 in CUB and TED, which show the greatest change between C3 and C3b, correspond to those identified by NMR¹¹. Relocation of TED in C3b positions its cell-surface-binding region opposite the β -ring domains⁵, placing a string of proposed binding sites for protein ligands in an array between the β -ring and the cell surface (Fig. 3a).

b, Largest changes. Left: positions of CUBg and CUBf in C3b (pink and blue), C3 (brown and cyan) and TED (red, C3; green, C3b); displacement of TED is visible. Right: top, changes in C345C (blue) and ANC (magenta); bottom, MG8 movement, from C3 (magenta) to C3b (blue).

One such important ligand is factor B, a serine protease, which initiates the formation of AP C3 convertase (C3bBb) by associating with activator surface-bound C3b. The resulting C3bBb is activated through cleavage by factor D protease and is stabilized by the binding of properdin¹². The stabilized convertase generates numerous C3b molecules, both by self-amplification and by the amplification of CP/LP through C3 convertase, which also generate C3b, which could form additional AP convertase complexes. Antibody and peptide binding¹³ and site-directed mutagenesis¹⁴ indicated a possible factor-B-binding site within residues 730–739 in the ANT domain of C3b. Mutation of 730DE and 736EE significantly decreased the

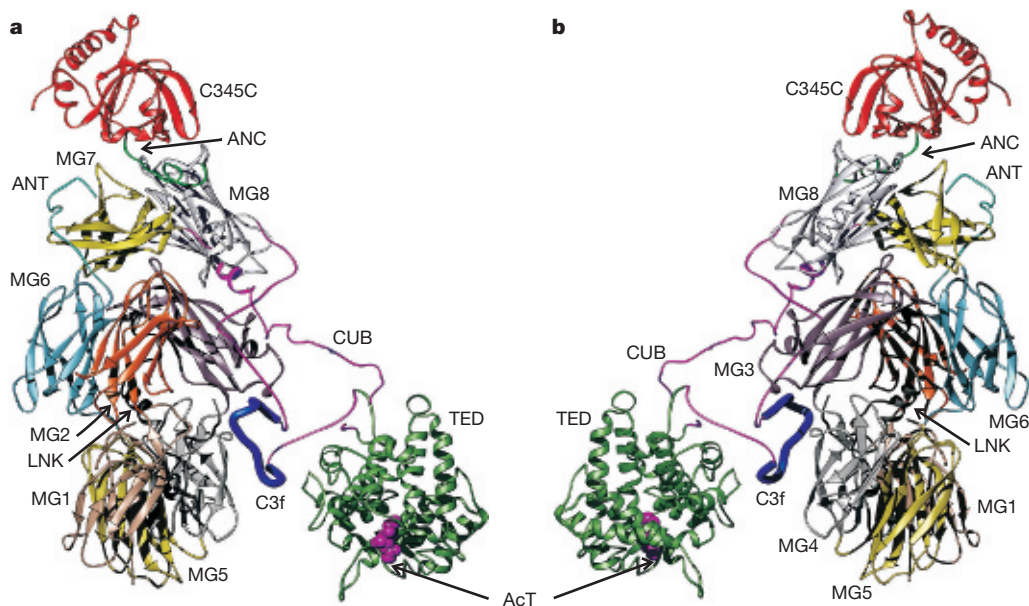


Figure 2 | Overall view of C3b. **a**, Each domain is rendered in a different colour and labelled. **b**, View rotated through 180°. A large separation of the TED domain from the β -ring is evident. The C3f region of CUBf (residues

1281–1299) is shown as a thick blue worm. The membrane attachment site, labelled AcT, is shown as magenta spheres and indicated by arrows.

binding of C3b to factor B (ref. 14). ANT, nearly completely buried in C3, switches to the surface of C3b (Fig. 3b). This motion, combined with movements of MG7 and MG8, makes residues 730DE and 736EE, buried in C3, fully accessible in C3b (Fig. 3c). However, the dispositions of ANT in C3b and C3c are essentially identical (Fig. 1a), although C3c does not bind factor B or support convertase assembly. Therefore, although ANT may provide one site for factor B binding, additional sites absent or obstructed on C3c must exist on C3b. Studies on a cobra-venom cleavage fragment of C3 (C3o) indicate that C3o might contact factor B through residues 933–942 in CUBg¹⁵. Unlike C3c, C3o lacks residues 727–736 but retains the 933–942

region, and supports the formation of the AP C3 convertase. Residues 933–942 are in the middle of CUBg, the unravelling of which exposes them to ligand binding in C3b (Fig. 3a). These observations indicate that factor B might contact both ANT and parts of CUBg. In addition, studies with an anti-C3d antibody indicate that factor B might also interact with C3d, emphasizing the complex nature of interactions between C3b and factor B (ref. 16). Structure-guided mutational studies are apparently needed to describe the topology of factor-B-binding sites on C3b more accurately. The AP convertase might be stabilized by tripartite interactions between C3b, properdin and Bb¹⁷, with C3b interacting with

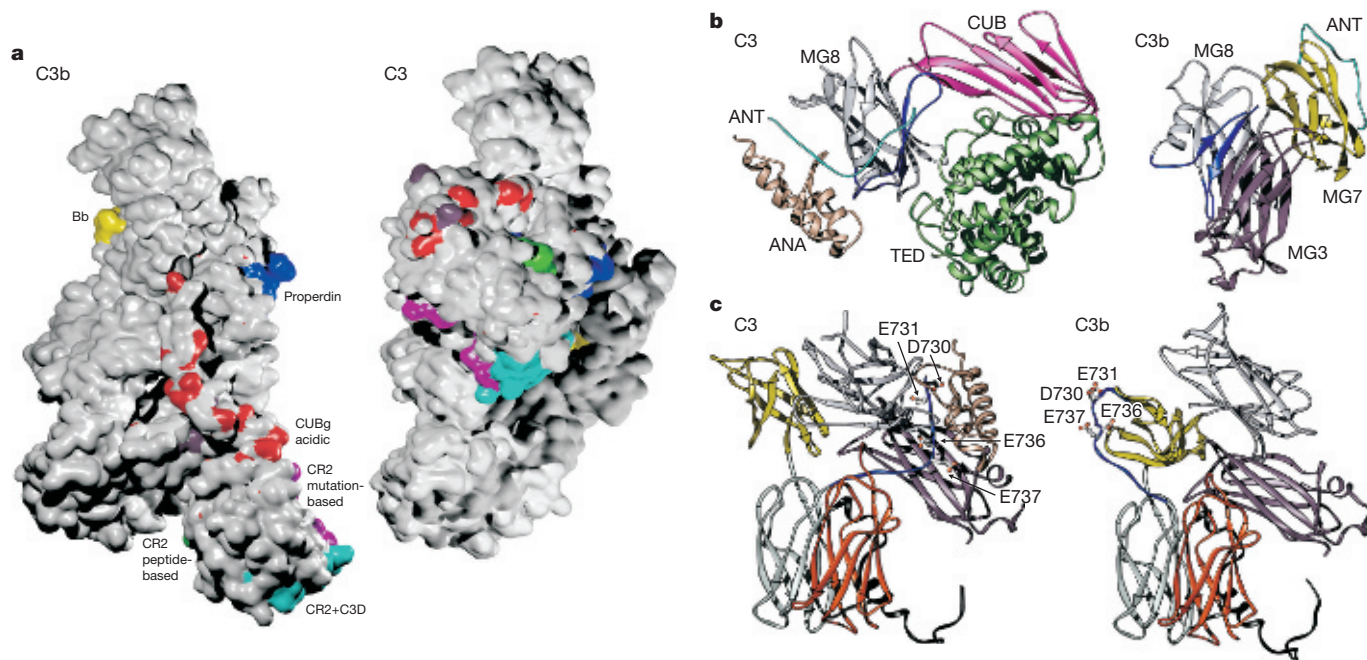


Figure 3 | Potential binding sites. **a**, Orientations of binding sites on C3b (left) and C3 (right). Solvent-accessible surfaces are coloured grey. Surface patches formed by residues shown experimentally to contact different ligands are shown in different colours and labelled. **b**, ANT (cyan) in C3b (right) shows a positional shift from C3 (left), making it more accessible.

Residues 1403–1435 (blue), a potential P-binding site, become more accessible in C3b as a result of a combination of excision of ANA (gold in C3) and displacement of ANT. **c**, Exposure of 730DE and 736EE (sticks) in C3b-ANT (blue, right) from their locations in C3 (left).

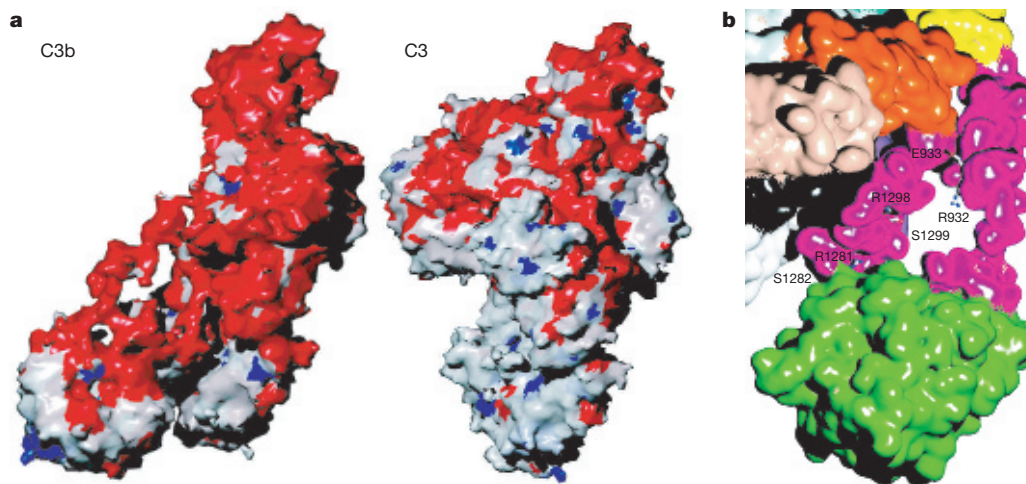


Figure 4 | C3b inactivation. **a**, Electrostatic surface charge on C3b (left) and C3, contoured at $\pm 5 \text{ e} \text{ \AA}^{-2}$. The C3b surface is significantly more negative. **b**, Two Arg-Ser bonds and an Arg-Glu peptide bond that are cleaved in converting C3b to C3c are shown as sticks. Solvent-accessible surfaces of the

β -ring domains, coloured white, yellow, orange and beige, and TED (green) and CUBg and CUBf (magenta) are also shown. Sequestration of scissile bonds in a cavity that shields them from access to large proteases such as factor I is evident.

properdin through residues 1403–1435 (ref. 18). A combination of removal of ANA and transformations in TED and MG8 exposes the pair of β -strands formed by 1403–1435 (in MG8) to ligands (Fig. 3b). Although accompanied by only a small change in solvent accessibility ($184 \text{ \AA}^2$), its repositioning could eliminate barriers for ligand binding.

Our structure of C3b also provides insight into regulatory events in complement function. Complement activation is regulated largely through manipulation of the stability of C3 convertase and C5 convertase. In blood, AP C3 convertase and C5 convertase are rapidly dissociated, and their C3b component is proteolysed by factor I, a serine protease. Both dissociation and proteolysis are mediated by regulatory proteins through their decay acceleration and factor I cofactor activities, respectively. Regulatory proteins are largely made up of repeating units, each about 60 residues long, called short consensus repeats or complement control protein modules¹⁹. Both decay acceleration and factor I cofactor activities involve the specific binding of regulatory proteins to C3b, with a strong dependence on ionic strength. Structural studies²⁰ and site-specific mutagenesis²¹ have indicated that positive charges on regulators, and negative charges on C3b, might have a key function in this interaction. It is therefore intriguing to observe that C3b has a larger negative surface-charge density than C3 (Fig. 4a). Negative charge augmentation is partly a result of an increase in $1,230 \text{ \AA}^2$ of accessible surface area for acidic side chains (Asp, Glu and Tyr) of CUB and TED domains in C3b in comparison with C3. Exposed acidic residues might also provide additional anchoring points for regulators. Most regulators bind to multiple clusters of acidic residues on C3b. For example, factor H recognizes two such clusters on C3b: on residues 744–754 in ANT²² and on residues 1187–1249 in TED²³. The distance between these is about $99 \text{ \AA}$ in C3b. Electron microscopy and solution scattering²⁴ data indicate a separation of 85 – $295 \text{ \AA}$ for C3b-binding sites in factor H. Thus, the size and flexibility of factor H could easily enable it to span these sites separated by $99 \text{ \AA}$. Vaccinia virus complement-control protein (VCP) binds to C3dg (residues 933–1281)²⁵ but not to C3d, indicating that the presence of acidic clusters in CUBg (Fig. 3a) might provide anchoring points for positively charged areas in VCP^{20,26}. Cofactor activity depends on the formation of a C3b–cofactor–factor-I complex, followed by conformational changes that permit the cleavage of C3b by factor I (ref. 1). Three peptide bonds, Arg 1281–Ser 1282, Arg 1298–Ser 1299 and Arg 932–Glu 933, in C3b are cleaved successively, resulting in a complete loss of the ability to form a C3 convertase¹. Rearrangement of the CUB domain results in the sequestration of all three scissile bonds in a cavity formed by the β -ring, TED and CUBg (Fig. 4b). Shielding of the scissile bonds in this cavity

hinders direct access by factor I, a large two-chain glycoprotein²⁷. Our structure of C3b therefore shows the necessity for a conformational change in C3b to enable factor I to access the scissile bonds. Moreover, the surface of C3b surrounding the scissile bonds is largely negatively charged (Fig. 4a), making the approach and interaction with factor I, which has a negatively charged Asp 90 at its substrate-binding site, unfavourable²⁷. Thus, in addition to facilitating conformational changes required for hydrolysis, positively charged cofactors might also electrostatically assist the approach and docking of C3b with factor I by reducing the negative charge on C3b. Structural characterization of this central component of complement is likely to provide a basis for a better understanding of complement function and for the potential therapy of complement-related diseases.

METHODS

Crystallization, structure determination and refinement. C3b was prepared by trypsin cleavage of C3 and purified as described²⁸. The cleavage was performed in the presence of 10 mM AcT to provide a nucleophile for covalent attachment of the side-chain carbonyl of Gln 991, on activation of the internal thioester³. Esterification was confirmed through mass spectroscopic analysis on tryptic digests of modified C3b. Modified protein was stored at a concentration of 7.7 mg ml^{-1} in 20 mM Tris-HCl pH 7.5, and a protein inhibitor cocktail. Crystals were obtained through vapour diffusion from $2 \mu\text{l}$ of the protein solution mixed with an equal volume of well solution. Wells contained 200 mM Tris-HCl pH 7.5, 100 mM NaCl, 20 mM LiCl and 15% PEG6000. Although crystals could also be grown with C3b prepared without AcT, they did not diffract beyond $6.5 \text{ \AA}$. AcT modification improved diffraction to $2.3 \text{ \AA}$. Crystals had a solvent content of 79% and were mechanically fragile, but reflection data set to $2.3 \text{ \AA}$ could be assembled from partial measurements on four crystals. Data were measured on flash-frozen crystals at 100 K on the ID22 beamline (Advanced Photon Source) at $0.9791 \text{ \AA}$. Statistics for data processing with the HKL2000 program package (ref. 29) are given in Supplementary Table 1. The structure was determined through molecular replacement with Phaser³⁰, by using fragments of the structures of C3c and C3 (ref. 5) as search models. The β -ring of C3c was first located; remaining domains of C3c and TED domain from C3 were then added sequentially, alternating with rigid-body refinement. CUB domains were built into difference maps phased on the rest of the structure. The structure of C3b, consisting of $1,564$ residues, has been refined in REFMAC³⁰ to an R value of 18.0 with a free R value of 19.4 . Representative geometric parameters are summarized in Supplementary Table 1 and an overall view of the structure is shown in Fig. 2. There are 20 residues in disallowed regions of the Ramachandran map, and fewer *cis* peptides than in C3 (ref. 5).

Received 20 July; accepted 18 September 2006.

Published online 15 October 2006.

- Volanakis, J. E. & Frank, M. M. *The Human Complement System in Health and Disease* (Marcel Dekker, New York, 1998).

2. Lambris, J. D. The multifunctional role of C3, the third component of complement. *Immunol. Today* **9**, 387–393 (1988).
3. Law, A. S. K. & Dodds, A. W. The internal thioester and the covalent binding properties of the complement proteins C3 and C4. *Protein Sci.* **6**, 263–274 (1997).
4. Kotwal, G. J. Poxviral mimicry of complement and chemokine system components: what's the end game? *Immunol. Today* **21**, 242–248 (2000).
5. Janssen, B. J. *et al.* Structures of complement component C3 provide insights into the function and evolution of immunity. *Nature* **437**, 505–511 (2005).
6. Feinberg, H. *et al.* Crystal structure of the CUB1-EGF-CUB2 region of mannose-binding protein associated serine protease-2. *EMBO J.* **22**, 2348–2359 (2003).
7. Nagar, B., Jones, R. G., Diefenbach, R. J., Isenman, D. E. & Rini, J. M. X-ray crystal structure of C3d: a C3 fragment and ligand for complement receptor 2. *Science* **280**, 1277–1281 (1998).
8. Zanotti, G. *et al.* Structure at 1.44 Å resolution of an N-terminally truncated form of the rat serum complement C3d fragment. *Biochim. Biophys. Acta* **1478**, 232–238 (2000).
9. Puliti, R. & Mattia, C. A. Ammonium N-acetyl-L-threoninate and methylammonium N-acetyl-L-threoninate. *Acta Crystallogr. C* **56**, 496–499 (2000).
10. Gettins, P. G. Thiol ester cleavage-dependent conformational change in human α 2-macroglobulin. Influence of attacking nucleophile and of Cys949 modification. *Biochemistry* **34**, 12233–12240 (1995).
11. Winters, M. S., Spellman, D. S. & Lambris, J. D. Solvent accessibility of native and hydrolyzed human complement protein 3 analyzed by hydrogen/deuterium exchange and mass spectrometry. *J. Immunol.* **174**, 3469–3474 (2005).
12. Xu, Y., Narayana, S. V. & Volanakis, J. E. Structural biology of the alternative pathway convertase. *Immunol. Rev.* **180**, 123–135 (2001).
13. Becherer, J. D., Alsenz, J., Esparza, I., Hack, C. E. & Lambris, J. D. Segment spanning residues 727–768 of the complement C3 sequence contains a neoantigenic site and accommodates the binding of CR1, factor H, and factor B. *Biochemistry* **31**, 1787–1794 (1992).
14. Clemenza, L. & Isenman, D. E. Structure-guided identification of C3d residues essential for its binding to complement receptor 2 (CD21). *J. Immunol.* **165**, 3839–3848 (2000).
15. O'Keefe, M. C., Caporale, L. H. & Vogel, C. W. A novel cleavage product of human complement component C3 with structural and functional properties of cobra venom factor. *J. Biol. Chem.* **263**, 12690–12697 (1988).
16. Koistinen, V., Wessberg, S. & Leikola, J. Common binding region of complement factors B, H and CR1 on C3b revealed by monoclonal anti-C3d. *Complement Inflamm.* **6**, 270–280 (1989).
17. Higgins, J. M., Wiedemann, H., Timpl, R. & Reid, K. B. Characterization of mutant forms of recombinant human properdin lacking single thrombospondin type I repeats. Identification of modules important for function. *J. Immunol.* **155**, 5777–5785 (1995).
18. Daoudaki, M. E., Becherer, J. D. & Lambris, J. D. A 34-amino acid peptide of the third component of complement mediates properdin binding. *J. Immunol.* **140**, 1577–1580 (1988).
19. Bork, P., Downing, A. K., Kieffer, B. & Campbell, I. D. Structure and distribution of modules in extracellular proteins. *Q. Rev. Biophys.* **29**, 119–167 (1995).
20. Krishna Murthy, H. M. *et al.* Crystal structure of a complement control protein that regulates both pathways of complement activation and binds heparan sulfate proteoglycans. *Cell* **104**, 301–311 (2001).
21. Smith, B. O. *et al.* Structure of the C3b binding site of CR1 (CD35), the immune adherence receptor. *Cell* **108**, 769–780 (2002).
22. Oran, A. E. & Isenman, D. E. Identification of residues within the 727–767 segment of human complement component C3 important for its interaction with factor H and with complement receptor 1 (CR1, CD35). *J. Biol. Chem.* **274**, 5120–5130 (1999).
23. Lambris, J. D., Avila, D., Becherer, J. D. & Muller-Eberhard, H. J. A discontinuous factor H binding site in the third component of complement as delineated by synthetic peptides. *J. Biol. Chem.* **263**, 12147–12150 (1988).
24. Aslam, M. & Perkins, S. J. Folded-back solution structure of monomeric factor H of human complement by synchrotron X-ray and neutron scattering, analytical ultracentrifugation and constrained molecular modelling. *J. Mol. Biol.* **309**, 1117–1138 (2001).
25. Bernet, J., Mullick, J., Panse, Y., Parab, P. B. & Sahu, A. Kinetic analysis of the interactions between vaccinia virus complement control protein and human complement proteins C3b and C4b. *J. Virol.* **78**, 9446–9457 (2004).
26. Ganesh, V. K., Smith, S. A., Kotwal, G. J. & Murthy, K. H. Structure of vaccinia complement protein in complex with heparin and potential implications for complement regulation. *Proc. Natl Acad. Sci. USA* **101**, 8924–8929 (2004).
27. Tsiftoglou, S. A. *et al.* The catalytically active serine protease domain of human complement factor I. *Biochemistry* **44**, 6239–6249 (2005).
28. Smith, S. A. *et al.* Mapping of regions within the vaccinia virus complement control protein involved in dose-dependent binding to key complement components and heparin using surface plasmon resonance. *Biochim. Biophys. Acta* **1650**, 30–39 (2003).
29. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276A**, 307–326 (1997).
30. CCP4. The CCP4 Suite: Programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Chrzas for help with data measurement, K. Judge for technical assistance, the UAB Mass Spectroscopy Facility for mass spectra, and L. DeLucas for support. This research was supported by the NIH (H.M.K.M. and S.V.L.N.). G.J.K. is an International Senior Wellcome Trust Fellow for Biomedical Sciences in South Africa.

Author Contributions K.G. and H.M.K.M. performed the crystallization and data measurement; A.A.A. and H.M.K.M. conducted the structure solution and refinement, and A.A.A., H.M.K.M., J.E.V., S.V.L.N. and G.J.K. performed the analysis.

Author Information Coordinates and structure factors are deposited in the Protein Data Bank under accession number 2HR0. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.M.K.M. (murthy@cbse.uab.edu).

LETTERS

Wza the translocon for *E. coli* capsular polysaccharides defines a new class of membrane protein

Changjiang Dong^{1*}, Konstantinos Beis^{1*†}, Jutta Nesper^{2†}, Anne L. Brunkan-LaMontagne², Bradley R. Clarke², Chris Whitfield² & James H. Naismith¹

Many types of bacteria produce extracellular polysaccharides (EPSs). Some are secreted polymers and show only limited association with the cell surface, whereas others are firmly attached to the cell surface and form a discrete structural layer, the capsule, which envelopes the cell and allows the bacteria to evade or counteract the host immune system¹. EPSs have critical roles in bacterial colonization of surfaces², such as epithelia and medical implants; in addition some EPSs have important industrial and biomedical applications in their own right³. Here we describe the 2.26 Å resolution structure of the 340 kDa octamer of Wza, an integral outer membrane lipoprotein, which is essential for group 1 capsule export in *Escherichia coli*. The transmembrane region is a novel α -helical barrel. The bulk of the Wza structure is located in the periplasm and comprises three novel domains forming a large central cavity. Wza is open to the extracellular environment but closed to the periplasm. We propose a route and mechanism for translocation of the capsular polysaccharide. This work may provide insight into the export of other large polar molecules such as DNA and proteins.

There are two distinct capsule assembly pathways in *E. coli*⁴. The defining characteristic of the group 1 (Wzy-dependent) pathway is that individual lipid-linked polymer repeat units are synthesized and exported to the periplasm, where a putative polymerase (Wzy) assembles the full length polymer⁴. *E. coli* capsular serotype K30 is the prototype for the group 1 capsule assembly system and the working model of the pathway is shown in Fig. 1a. The K30 polysaccharide is assembled and exported by proteins encoded by a 12-gene operon⁵. The process requires a member of the OMA (outer membrane auxiliary) protein family for translocation of nascent polymer across the outer membrane⁶. Mature Wza, the best studied OMA member, is a 359-residue lipoprotein protein that forms SDS-stable octamers; it is synthesized as a precursor with a cleavable 20-residue amino-terminal signal sequence⁶. Cys 21 is modified by a thioether-linked diacylglyceryl group and its amino group is acylated. Electron microscopy studies of two-dimensional crystals of Wza protein in lipid bilayers revealed octameric ring-like structures suggestive of a channel⁷. In negatively stained cryo-electron-microscopy (cryo-EM) images (15.5 Å resolution), Wza octamers form a mushroom-like structure with dimensions 90 Å × 90 Å × 100 Å (ref. 8) with a large central cavity, but these images provided no mechanism for translocation. Chromosomal 'knockout' *wza* mutants have no detectable capsule and produce no intracellular polymer^{6,7}. This suggests a feedback process in which synthesis and export are coupled. Wza

interacts with a tetrameric inner membrane tyrosine autokinase protein, Wzc^{7,9,10}. Mutations that either eliminate Wzc, or compromise its phosphorylation, also turn off capsular polymer formation^{11,12}. A non-acylated Wza mutant forms weakly stable oligomers, and polymer accumulates in the periplasm⁷. This may indicate that the non-acylated oligomer is not competent for export but maintains its interaction with Wzc, thus promoting polymer synthesis. Sequence relationships amongst OMA proteins are limited⁶, but family members contain the polysaccharide biosynthesis/export (PES) motif (Pfam02563) (Supplementary Figs 1, 2).

The 2.26 Å crystal structure of mature acylated Wza from *E. coli* K30 has eight monomers (Fig. 2a) in the asymmetric unit and exhibits eightfold rotational symmetry (Fig. 2b, c). We describe the octameric structure as having the shape of an 'amphora' without the handles (Fig. 2b). Wza has a large internal cavity that is open at a narrow 'neck' in the outer membrane and closed at its base. The long axis of the molecule is approximately 140 Å and the diameter at the widest point is 105 Å (Fig. 2b, c). Comparison of the crystal structure with that derived from negatively stained cryo-EM⁸ reveals that the 'neck' is missing in the cryo-EM reconstruction. This region may be unstructured in the conditions of the cryo-EM experiment. The dimensions of the remainder of the structure and the large central cavity match well. The cryo-EM structure has only fourfold rotational symmetry indicating conformational change may be induced by negative staining⁸; images of Wza in proteolipid samples showed eightfold symmetry⁷.

Domain 1 of Wza (residues 89–169) represents a novel fold; it comprises an anti-parallel β -sandwich with an α -helix at one edge. Domain 1 contains the conserved PES motif (Supplementary Figs 1b, 2). The eight copies of domain 1 combine to form a ring structure (ring 1) at the bottom of the Wza, with an eightfold axis through the centre (Fig. 2b, c). Ring 1 presents a concave surface at the base of the structure and the centre is filled by eight loops (residues 105 to 112 of each domain). Tyr 110 is located at the tip of each loop but is not clearly visible in the experimental map, suggesting that the side chain is flexible (Supplementary Fig. 3a). Inwards past Tyr 110, the ring has an internal diameter of over 25 Å (compared with 17 Å at the base of ring 1; Fig. 2b, c). Domain 2 (residues 68–84 and 175–252) has a novel structure, despite possessing similar dimensions to domain 1. This domain has a central five-stranded mixed β -sheet with three α -helices on one face. The eight copies of domain 2 form an eightfold symmetric ring structure (ring 2) with an inner diameter of 25 Å. Domain 3 (residues 46–64 and 255–344) is a structural duplication

¹Centre for Biomolecular Sciences, The University of St Andrews, Fife KY16 9RH, UK. ²Department of Molecular and Cellular Biology, The University of Guelph, Guelph, Ontario N1G 2W1, Canada. [†]Present addresses: Department of Biochemistry, Imperial College of Science, Faculty of Life Sciences, Division of Molecular Biosciences, Wolfson Building, London SW7 2AZ, UK (K.B.); Department of Biology, University of Konstanz, 78457 Konstanz, Germany (J.N.).

*These authors contributed equally to this work.

of domain 2 (Supplementary Fig. 1c) and the eight copies come together to form a ring structure (ring 3; Fig. 2b) with a significantly larger external diameter (105 Å) than ring 2. The three rings sit one

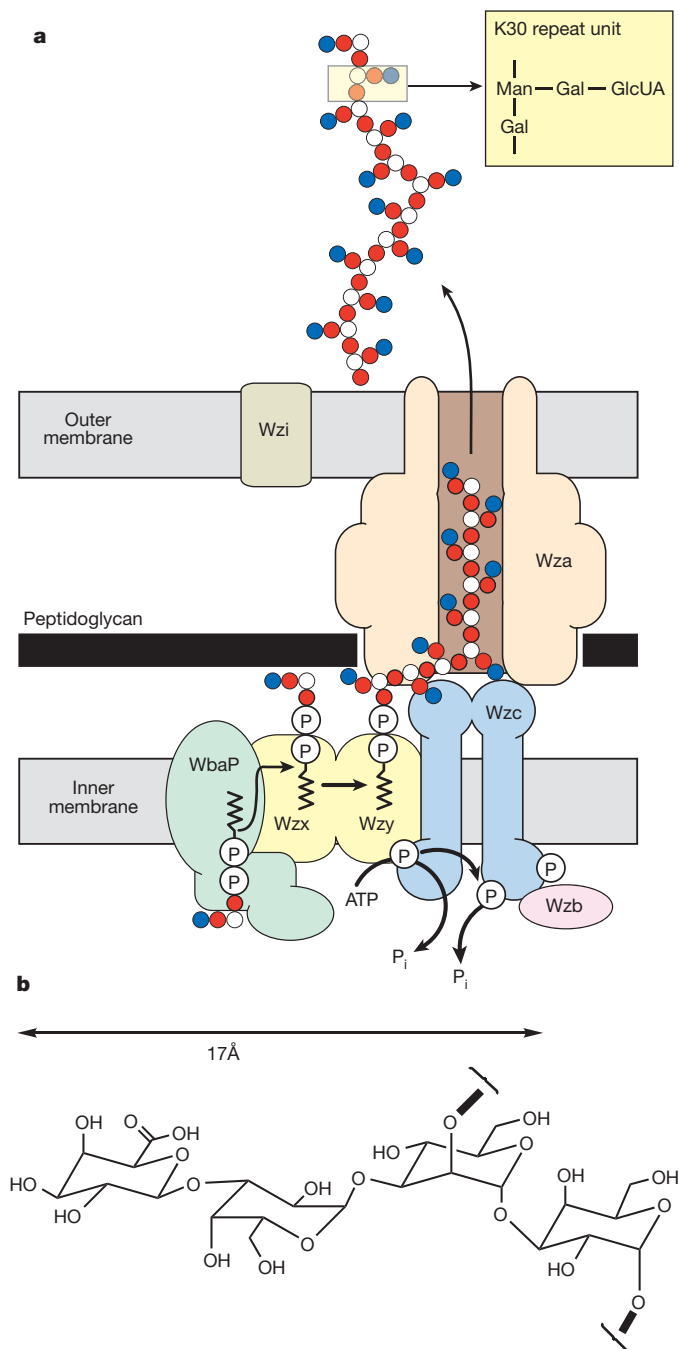


Figure 1 | Group 1 capsular polysaccharide export in Gram-negative bacteria. **a**, Model of the biosynthetic complex carrying out synthesis and export of serotype K30 group 1 capsule in *E. coli*⁶. Repeat units of the polymer are assembled on a lipid (undecaprenol diphosphate; und-PP) acceptor, in the cytoplasm in a reaction initiated by the WbaP enzyme. The und-PP-linked repeat units are 'flipped' across the inner membrane by Wzx. Polymerization occurs at the periplasmic face and is dependent on another integral membrane protein, the putative polymerase, Wzy. Wzy-dependent polymerization requires the activity of the tetrameric Wzc protein¹⁰. Wzb is a protein tyrosine phosphatase enzyme responsible for dephosphorylating Wzc. The cycling phosphorylation of Wzc is crucial for export^{11,12}. Wza and Wzc proteins interact to form a complex that spans the periplasm^{6,7}. The precise role of Wzi, an integral membrane protein, is unclear. **b**, The chemical structure of the repeat unit has a width of 17 Å (assuming the most extended conformation).

atop the other with a spacing of 10 Å between rings 2 and 3. A ribbon representation of the structure (Fig. 2b) gives the appearance of side holes being present, but these are filled by side chains (Fig. 3a, b). Domain 4 comprises the carboxy terminus of the monomer (residues 345–376; Fig. 2a) and is an amphipathic helix. As a result of the rotational symmetry, the helices from the eight monomers create an α -helical barrel (Fig. 2b) at the 'neck' of the molecule. The barrel is tapered such that the end proximal to ring 3 has an internal diameter of approximately 30 Å. At the open end the internal diameter is 17 Å. The N terminus (residues 22–45) is wrapped around the top of ring 3 (Fig. 2a, b).

The central cavity of the protein has an internal volume of approximately 15,000 Å³ (Figs 2b, 3a), comparable to GroEL–GroES (84,000 Å³; ref. 13) and TolC (44,000 Å³; ref. 14). In Wza, the protein surface that encloses the central cavity is polar (Fig. 3a) and lined by residues with little sequence conservation. Those conserved residues that do exist are structural or located at the extensive subunit interfaces (Supplementary Information and Supplementary Fig. 2). The exterior surface of Wza is almost all polar (Fig. 3b). The exception is the helical barrel which has a markedly hydrophobic exterior (Fig. 3b) with bound lipid molecules and Trp 350 exposed (Fig. 3b and Supplementary Figs 3b, 4). The helical barrel is the transmembrane region; but the bulk of the structure is located in the periplasm. The C

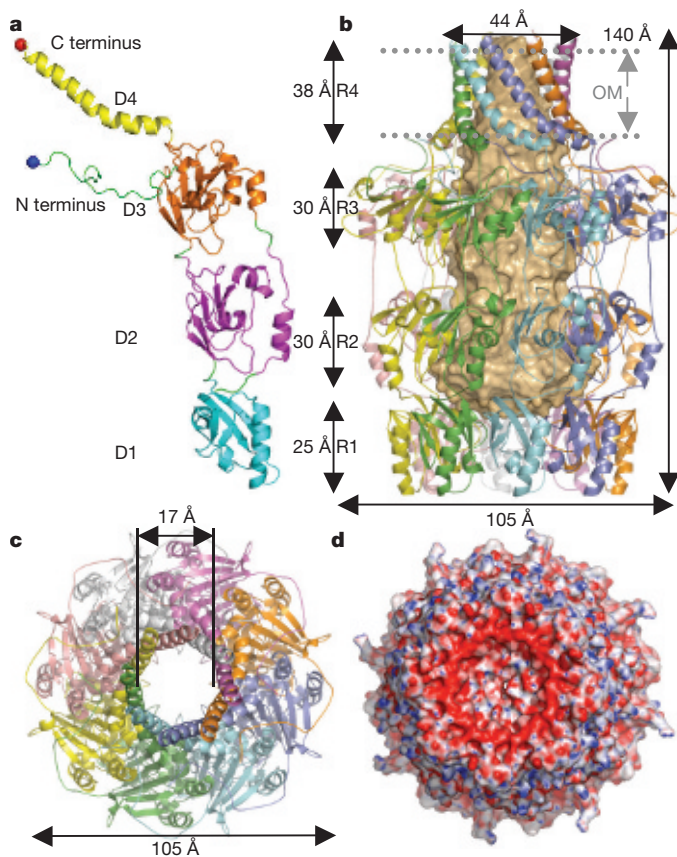


Figure 2 | The structure of Wza. **a**, Wza has four domains labelled as D1–4. Cys 21 (acylated N terminus) is shown as a blue ball and Arg 376 (last ordered residue) as a red ball. **b**, The Wza octamer is described as an amphora made of four rings (R1–4) with a large central cavity (orange space-filling shape). Ring 1 is formed by eight copies of domain 1, ring 2 by eight copies of domain 2, and so on. The helical barrel (R4) forms the 'neck' and ring 1 the 'base'. The outer membrane (OM) is marked. The protein C terminus is outside the cell and rings 1, 2 and 3 inside the periplasm. **c**, View into the central cavity from outside the cell. **d**, View of the central cavity from the periplasm with Wza (surface render) coloured by electrostatic charge (blue, positive; red, negative). The concave surface of ring 1 is closed by each loop at Tyr 110 and has a band of negative charge.

terminus of each monomer is exposed on the cell surface (Fig. 2b), placing the acylated N terminus at the inner leaflet of the outer membrane, consistent with typical outer membrane lipoproteins¹⁵. The orientation of Wza was established by showing that a fused C-terminal Flag epitope (Wza-Flag) is exposed on the cell surface (Fig. 3c and Supplementary Fig. 5). To date, all integral outer membrane proteins have contained a transmembrane β -barrel¹⁶. Wza is the first example of a transmembrane α -helical barrel. Although transmembrane α -helices are well known in bacterial inner membrane proteins and eukaryotic membrane proteins, the integral membrane helical barrel arrangement seen in Wza has not been observed before. Channel-forming antimicrobial peptides have amphipathic helices¹⁷ and the C terminus of Wza may mimic antimicrobial peptide pore formation.

A simple model for translocation has the carbohydrate moving from the periplasm to the central cavity of Wza and exiting through the helical barrel. Assuming an extended conformation, the width of the oligosaccharide is approximately 17 Å (Fig. 1b), matching the internal diameter of the helical barrel (Figs 2c, 3a). Unlike β -barrel porin proteins that are involved in the uptake of small nutrients¹⁸, Wza has no portal between the central cavity and periplasm (Fig. 2d and Supplementary Fig. 3c, d). Isolated Wza protein embedded in lipid bilayers shows no ion conductance (C. Whitfield and R. E. W. Hancock, unpublished data) consistent with a closed state. Regulation of the opening of the Wza central cavity would seem desirable in maintaining outer membrane barrier properties. We propose that the opening of the portal requires a substantial conformational change. Both Wza^{Y110A} and Wza^{Y110W} derivatives of Wza are fully competent for capsule export (Supplementary Fig. 6) indicating that gating is not controlled by a single residue. The trigger for such hypothetical conformation changes is unknown but the binding of Wzc (refs 7, 9) provides one logical candidate. The circle

of negative charge and a concave surface at the base of the Wza oligomer, the conserved PES domain, offers a potential site for protein–protein interactions (Fig. 2d). The use of conformational changes to gate large channels was observed in FhuA (ref. 19). TolC has a barrel closed at one end that opens in response to periplasmic binding events¹⁴. Wza may represent a model of some of the multimeric ‘secretin’ protein complexes found in type II, III and IV secretion systems, most notably single-stranded DNA export through type IV systems²⁰. Electron microscopy structures of several secretins show these contain ring-like structures capable of accommodating large polar substrates^{21–23}.

The potassium transporter²⁴ and the conceptually similar ammonia transporter²⁵ use side chains inside the channel to bind substrate and thus avoid the desolvation penalty^{24,25}. Carbohydrates contain many polar atoms whose desolvation *en masse* would be unfavourable. Lactose permease binds the two carbohydrate rings inside a central cavity²⁶ satisfying the hydrogen bonds. The protein then undergoes a conformational change, releasing the sugar on the opposite side of the membrane. A similar model has been proposed for the lipopolysaccharide lipid A flippase, MsbA²⁷. Extended carbohydrate binding sites in proteins are well known²⁸, but are characterized by exquisite specificity. Using protein side chains to recognize a transient EPS chain with hundreds of carbohydrate rings, would seem a formidable task. We propose that Wza, by permitting both water molecules and carbohydrate to be present simultaneously in its large polar cavity, uses water molecules to ‘plug’ gaps. This ensures that the polar protein side chains and sugar hydroxyl groups make hydrogen bonds. This ‘lubrication’ of export negates the need for specific recognition between protein and carbohydrate. In support of this model, we note that an essentially identical Wza is employed by *E. coli* and *Klebsiella pneumoniae*²⁹ despite them having chemically distinct EPSs. Moreover, Wza homologues that export different EPSs can complement Wza-deficient *E. coli* K30 (refs 6, 9).

METHODS

Full details of the purification and crystallization of the native protein have been published³⁰. The 2.26 Å data set was collected on ID29 of the European Synchrotron Radiation Facility (ESRF) on a single selenomethionine crystal protected with 25% (w/v) glycerol in the mother liquid and cooled to 100 K. Experimental phases were calculated by locating Se atoms. A full description of the structural analysis is given in Supplementary Information.

Plasmid pWQ126 (ref. 6) provided the template for mutagenesis using the QuikChange Site-Directed Mutagenesis Kit (Stratagene). Y110A, Y110W and Δ 108–112 (deletion) mutants were constructed. The phenotypes of the mutants were determined after transforming the wza-null strain *E. coli* CWG281 (*wza*_{K30}::*aacC1*, *wza*_{22min}::*aadA*; Gm^r, Sp^r)⁶. The gene encoding the Flag-epitope at the C-terminus of the marker protein fusion (Wza-Flag) was constructed by PCR. Cells were grown and imaged as described in Supplementary Information.

Received 7 August; accepted 21 September 2006.

Published online 1 November 2006.

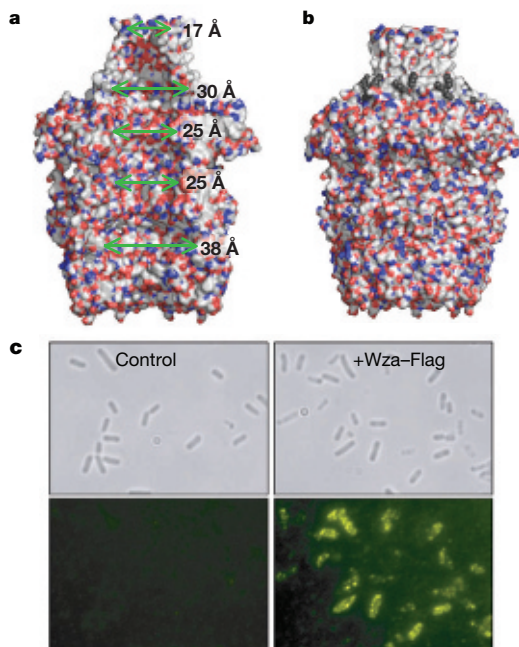


Figure 3 | The surfaces of Wza. **a**, The central cavity of Wza is large and polar (oxygen, red; nitrogen, blue; carbon, selenium and sulphur, white). **b**, Lipid molecules are shown as black spheres on the exterior surface. The helical barrel is clearly non-polar. A band of tryptophan residues is exposed on the helical barrel (Supplementary Fig. 4b). **c**, For both the control sample (*E. coli* LE392) and *E. coli* LE392 expressing Wza-Flag (+Wza-Flag), a differential interference contrast image is shown in the upper frame and the corresponding fluorescence image in the lower. The Flag-tag is clearly exposed on the cell surface.

1. Roberts, I. S. The biochemistry and genetics of capsular polysaccharide production in bacteria. *Annu. Rev. Microbiol.* **50**, 285–315 (1996).
2. Hall-Stoodley, L., Costerton, J. W. & Stoodley, P. Bacterial biofilms: from the natural environment to infectious diseases. *Nature Rev. Microbiol.* **2**, 95–108 (2004).
3. Sutherland, I. W. Novel and established applications of microbial polysaccharides. *Trends Biotechnol.* **16**, 41–46 (1998).
4. Whitfield, C. Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. *Annu. Rev. Biochem.* **75**, 39–68 (2006).
5. Drummelsmith, J. & Whitfield, C. Gene products required for surface expression of the capsular form of the group 1 K antigen in *Escherichia coli* (O9a:K30). *Mol. Microbiol.* **31**, 1321–1332 (1999).
6. Drummelsmith, J. & Whitfield, C. Translocation of group 1 capsular polysaccharide to the surface of *Escherichia coli* requires a multimeric complex in the outer membrane. *EMBO J.* **19**, 57–66 (2000).
7. Nesper, J. et al. Translocation of group 1 capsular polysaccharide in *Escherichia coli* serotype K30. Structural and functional analysis of the outer membrane lipoprotein Wza. *J. Biol. Chem.* **278**, 49763–49772 (2003).
8. Beis, K. et al. Three-dimensional structure of Wza, the protein required for translocation of group 1 capsular polysaccharide across the outer membrane of *Escherichia coli*. *J. Biol. Chem.* **279**, 28227–28232 (2004).

9. Reid, A. N. & Whitfield, C. Functional analysis of conserved gene products involved in assembly of *Escherichia coli* capsules and exopolysaccharides: evidence for molecular recognition between Wza and Wzc for colanic acid biosynthesis. *J. Bacteriol.* **187**, 5470–5481 (2005).
10. Collins, R. F. *et al.* Periplasmic protein–protein contacts in the inner membrane protein Wzc form a tetrameric complex required for the assembly of *Escherichia coli* group 1 capsules. *J. Biol. Chem.* **281**, 2144–2150 (2006).
11. Wugeditsch, T. *et al.* Phosphorylation of Wzc, a tyrosine autokinase, is essential for assembly of group 1 capsular polysaccharides in *Escherichia coli*. *J. Biol. Chem.* **276**, 2361–2371 (2001).
12. Paiment, A., Hocking, J. & Whitfield, C. Impact of phosphorylation of specific residues in the tyrosine autokinase, Wzc, on its activity in assembly of group 1 capsules in *Escherichia coli*. *J. Bacteriol.* **184**, 6437–6447 (2002).
13. Xu, Z. H., Horwich, A. L. & Sigler, P. B. The crystal structure of the asymmetric GroEL–GroES–(ADP)(7) chaperonin complex. *Nature* **388**, 741–750 (1997).
14. Koronakis, V., Sharff, A., Koronakis, E., Luisi, B. & Hughes, C. Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature* **405**, 914–919 (2000).
15. Tokuda, H. & Matsuyama, S. Sorting of lipoproteins to the outer membrane in *E. coli*. *Biochim. Biophys. Acta* **1693**, 5–13 (2004).
16. Ruiz, N., Kahne, D. & Silhavy, T. J. Advances in understanding bacterial outer-membrane biogenesis. *Nature Rev. Microbiol.* **4**, 57–66 (2006).
17. Bulet, P., Stocklin, R. & Menin, L. Anti-microbial peptides: from invertebrates to vertebrates. *Immunol. Rev.* **198**, 169–184 (2004).
18. Cowan, S. W. *et al.* Crystal structures explain functional properties of two *E. coli* porins. *Nature* **358**, 727–733 (1992).
19. Ferguson, A. D., Hofmann, E., Coulton, J. W., Diederichs, K. & Welte, W. Siderophore-mediated iron transport: crystal structure of FhuA with bound lipopolysaccharide. *Science* **282**, 2215–2220 (1998).
20. Christie, P. J., Atmakuri, K., Krishnamoorthy, V., Jakubowski, S. & Cascales, E. Biogenesis, architecture, and function of bacterial type IV secretion systems. *Annu. Rev. Microbiol.* **59**, 451–485 (2005).
21. Linderth, N. A., Simon, M. N. & Russel, M. The filamentous phage pIV multimer visualized by scanning transmission electron microscopy. *Science* **278**, 1635–1638 (1997).
22. Nouwen, N. *et al.* Secretin PulD: association with pilot PulS, structure, and ion-conducting channel formation. *Proc. Natl Acad. Sci. USA* **96**, 8173–8177 (1999).
23. Collins, R. F. *et al.* Structure of the *Neisseria meningitidis* outer membrane PilQ secretin complex at 12 Å resolution. *J. Biol. Chem.* **279**, 39750–39756 (2004).
24. Zhou, Y. F., Morais-Cabral, J. H., Kaufman, A. & MacKinnon, R. Chemistry of ion coordination and hydration revealed by a K⁺ channel-Fab complex at 2.0 Å resolution. *Nature* **414**, 43–48 (2001).
25. Khademi, S. *et al.* Mechanism of ammonia transport by Amt/MEP/Rh: structure of AmtB at 1.35 Å. *Science* **305**, 1587–1594 (2004).
26. Abramson, J. *et al.* Structure and mechanism of the lactose permease of *Escherichia coli*. *Science* **301**, 610–615 (2003).
27. Reyes, C. L. & Chang, G. Structure of the ABC transporter MsbA in complex with ADP–vanadate and lipopolysaccharide. *Science* **308**, 1028–1031 (2005).
28. Moothoo, D. N. & Naismith, J. H. Concanavalin A distorts the β-GlcNAc-(1→2)-Man linkage of β-GlcNAc-(1→2)-α-Man-(1→3)-[β-GlcNAc-(1→2)-α-Man-(1→6)]-Man upon binding. *Glycobiology* **8**, 173–181 (1998).
29. Rahn, A., Drummelsmith, J. & Whitfield, C. Conserved organization in the *cps* gene clusters for expression of *Escherichia coli* group 1 K antigens: Relationship to the colanic acid biosynthesis locus and the *cps* genes from *Klebsiella pneumoniae*. *J. Bacteriol.* **181**, 2307–2313 (1999).
30. Beis, K., Nesper, J., Whitfield, C. & Naismith, J. H. Crystallization and preliminary X-ray diffraction analysis of Wza outer-membrane lipoprotein from *Escherichia coli* serotype O9a:K30. *Acta Crystallogr. D* **60**, 558–560 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements J.H.N. is a Biotechnology and Biology Sciences Research Council (BBSRC) Career Development Fellow, and acknowledges funding from a Wellcome Trust programme grant for biological aspects of the research. C.W. holds a Canada Research Chair and acknowledges funding from the Canadian Institutes of Health Research. The experimental structural biology was performed by the Scottish Structural Proteomics Facility, which is funded by the Scottish Higher Education Funding Council and the BBSRC. We acknowledge the use of ESRF beamlines and are grateful for assistance from D. Bourgeois and G. Leonard with data collection. We thank L. Cuthbertson for assistance with bioinformatic analyses, R. Clarke for fluorescence-activated cell sorting, and G. Taylor, M. White and B. Hunter for a critical review of the manuscript.

Author Contributions C.D. optimized crystals, collected data, and solved, refined and analysed the Wza structure. K.B. and J.N. grew the first crystals. K.B. devised and refined the SeMet protocol. A.L.B.-L. and B.R.C. made and analysed the site-directed mutants and Wza–Flag fusion, respectively. C.W. oversaw the project, analysed the data and wrote the paper. J.H.N. oversaw the project, collected X-ray data, solved and refined the structure, analysed the data and wrote the paper.

Author Information Coordinates and data are available from the Worldwide Protein Data Bank, accession code 2j58. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.H.N. (naismith@st-and.ac.uk).

LETTERS

A functional RNAi screen for regulators of receptor tyrosine kinase and ERK signalling

Adam Friedman¹ & Norbert Perrimon¹

Receptor tyrosine kinase (RTK) signalling through extracellular-signal-regulated kinases (ERKs) has pivotal roles during metazoan development, underlying processes as diverse as fate determination, differentiation, proliferation, survival, migration and growth. Abnormal RTK/ERK signalling has been extensively documented to contribute to developmental disorders and disease, most notably in oncogenic transformation by mutant RTKs¹ or downstream pathway components such as Ras and Raf². Although the core RTK/ERK signalling cassette has been characterized by decades of research using mammalian cell culture and forward genetic screens in model organisms, signal propagation through this pathway is probably regulated by a larger network of moderate, context-specific proteins. The genes encoding these proteins may not have been discovered through traditional screens owing, in particular, to the requirement for visible phenotypes. To obtain a global view of RTK/ERK signalling, we performed an unbiased, RNA interference (RNAi), genome-wide, high-throughput screen in *Drosophila* cells using a novel, quantitative, cellular assay monitoring ERK activation. Here we show that ERK pathway output integrates a wide array of conserved cellular processes. Further analysis of selected components—in multiple cell types with different RTK ligands and oncogenic stimuli—validates and classifies 331 pathway regulators. The relevance of these genes is highlighted by our isolation of a Ste20-like kinase and a PPM-family phosphatase that seem to regulate RTK/ERK signalling *in vivo* and in mammalian cells. Novel regulators that modulate specific pathway outputs may be selective targets for drug discovery.

Genome-wide RNAi screens can identify novel components of signal transduction pathways (for example, see ref. 3). The outputs from traditional transcriptional pathway reporters may be biased by cell type, integrate unknown additional pathways, and not linearly reflect endogenous signalling. In contrast, we developed a phospho-specific antibody reporter of proximal, endogenous pathway activity. Monoclonal antibodies recognizing the dually phosphorylated, active form of mammalian ERK1/2 (dpERK) have been characterized to report the dynamic activation of the single *Drosophila* ERK isoform, Rolled, downstream of multiple RTKs^{4,5}.

Drosophila haemocytes express the insulin receptor (InR), which activates both the Akt and ERK pathways when stimulated with insulin (Supplementary Fig. 1b; ref. 5). S2R+ cells expressing yellow fluorescent protein (YFP)-tagged Rolled were seeded in 384-well plates aliquoted with a genome-wide dsRNA collection targeting 20,420 predicted unique genes. Following knockdown, unstimulated or insulin-stimulated cells were stained with fluorescently-conjugated dpERK antibodies, and dpERK fluorescence was converted to ERK–YFP normalized Z-scores (see Methods). In assay validation tests, RNAi of known pathway components such as *PTP-ER*, *InR*, *Ras1* and *Ksr* had expected effects on ERK activation (Supplementary Fig. 1d).

We isolated 1,168 annotated genes (Supplementary Table 1 and Supplementary Fig. 2a, c), including the core RTK/ERK pathway in *Drosophila* (Supplementary Figs 1a, 2d). The assay significantly enriches for genes conserved in metazoans; 62% of the genes have a human orthologue, and 23% have a potential homologue implicated in disease (Supplementary Fig. 2e). A broad array of cellular functions influences ERK signalling output (Fig. 1a), whereas 56% of the hits have no annotated molecular function. We found enrichment for specific gene functions (Supplementary Table 2), including neurogenesis, cytoskeletal maintenance, morphogenesis and cell proliferation. On the basis of median Z-scores, we observed that trafficking genes negatively regulated RTK/ERK activation following stimulation, whereas components of the proteasome and ribosome were positive regulators (Fig. 1b). Finally, we found that 41 genes that regulate both ERK activation and cell morphology⁶ fall into multiple phenotypic categories (Supplementary Table 3), suggesting that the genes control several pathways regulating morphology.

In contrast to the core signalling cassette, other RTK/ERK pathway regulators may be cell-type-specific or required only following stimulus downstream of a particular RTK. For example, in PC12 cells, an NGF-induced, RAP1-mediated, sustained ERK signal leads to differentiation whereas EGF-induced, transient ERK activation leads to proliferation⁷. In *Drosophila*, the requirement for the Shc adaptor is restricted to particular RTKs (ref. 8). We classified our primary hits by secondary screening in two cell lines (S2R+ and Kc167) and under stimulation of distinct RTKs: InR, *Drosophila* EGFR (DER) and the PDGF/VEGF homologue receptor (PVR), which is also active at baseline, stimulated by insulin, secreted Spitz (sSpi) and PDGF- and VEGF-related factor isoforms (PVFs), respectively. Of 362 genes tested, 91% (331) had a significant effect under one or more stimuli (Fig. 2a; Supplementary Tables 4, 5). Because two-thirds of these have mammalian orthologues (Supplementary Table 6), many may also regulate ERK signalling in humans. In separate tests using non-overlapping dsRNAs, we found little contribution to the false-positive rate of off-target effects, probably owing to the filtering of our primary hit list for selection of dsRNAs without significant predicted off-targets (see Methods).

To categorize the activity of 331 genes across seven experimental conditions, we performed hierarchical phenotypic clustering (Fig. 2a). Genes that strongly affected ERK signalling regardless of cell line or RTK-activation (in the same way as known pathway components; Fig. 2b) may be novel components of the core signalling cassette or general permissive factors for ERK activation, such as *CG31302* (which encodes the orthologue of human RIM-binding protein 1/2) and *CG30387* (whose human orthologous protein ARMS is required for sustained ERK activation through Trk; ref. 9), suggesting that some of the identified genes have a broader role in RTK signalling. The *Argonaute1* (AGO1) knockdown may prevent biogenesis of miRNAs required for negative regulation of ERK signalling, such as *let-7* family

¹Department of Genetics, Howard Hughes Medical Institute, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA.

members¹⁰. Importantly, *CG30476* was recently discovered as a novel Cnk interactor modulating RTK activity during *Drosophila* development¹¹, illustrating the power of our approach to discover new, core signalling pathway components.

Other genes modulating ERK activation may be required only in specific contexts (Fig. 2b). InR had little effect under EGF stimulus, whereas PVR unexpectedly potentiated InR signalling but had possibly antagonistic effects after DER activation. Similar RTK specificity was observed for the conserved genes *Serotonin transporter* (*SerT*), *zinc finger homeodomain 1* (*zfh1*) and *CG3878*, the orthologue for which was isolated from a screen in mammalian cells for tumour suppressors¹². We also observed cell-line specificity for regulators such as *CG6701*, a putative transcription factor/helicase, and *Ras64B*, a Ras paralogue, potentially as a result of the relative expression in each cell line.

Particularly striking in cell-type specificity were components of the cell cycle machinery. Whereas cell cycle progression is a consequence of ERK activation^{13,14}, this effect can also be cell-type dependent¹³. Our results suggesting that cell cycle components constitute a feedback mechanism for regulating ERK signalling is consistent with cdc2-mediated Mos accumulation activating MAP kinase (MAPK) in *Xenopus*¹⁵. As has been shown for the Jak/STAT pathway in *Drosophila*¹⁶, cell cycle machinery regulating signal transduction may be direct rather than through cell cycle control.

We isolated many other genes that have been characterized as ERK effectors, such as components of the general transcription machinery¹⁷ (Supplementary Table 4). This is consistent with the synthetic multivulval (*synMuv*) negative regulators of vulva formation in *Caenorhabditis elegans*¹⁸ and isolation of general transcription factors from multiple *Drosophila* genetic screens for RTK/ERK regulators. More broadly, we found 82 genes that also seemed to be regulated by the pathway, as determined from two recent transcriptional profiling studies *in vivo*^{14,19} (Supplementary Table 7). Our results demonstrate that steady-state, feedback inhibition or amplification of ERK signalling following component depletion may be more extensive than is currently appreciated from studies of single genes. Just as the targets of RTK/ERK signalling are diverse, this pathway integrates a similar diversity of cellular processes in determining the precise level of output.

Signal specificity can arise by cross-talk with other pathways within the signalling network or at transcriptional targets. Although individual components of other canonical pathways were identified (for example, *cact*, *shn*, *fz2*), the JNK/MAPK and Akt/Tor pathways consistently negatively affected RTK/ERK signalling (Supplementary Fig. 4b, c). As AP-1 transcription factors *D-jun* (*Jra*) and *D-fos* (*kay*) are direct effectors of the JNK and ERK MAPKs¹⁷, their isolation as negative regulators suggests feedback on proximal pathway activation. Similarly, positive regulators of the Akt/Tor pathway, such as *Akt*, *Rheb* and *Tor*, negatively regulated ERK, whereas negative regulators such as *Pten* and *Tsc2* (also known as *gig*) positively regulated ERK activation. This is consistent with the recent demonstration that mammalian *Akt1* knockdown induces ERK activation²⁰ and Akt inhibits Raf through phosphorylation²¹. Although Akt phosphorylation sites in *Drosophila* Raf are conserved, the direction of ERK regulation by pathway components downstream of Akt is more consistent with negative feedback within the Akt pathway at the level of InR itself²², and, hence, this is possibly how Akt affects the ERK pathway.

Although mutations in or overexpression of multiple RTKs have been implicated in various cancers¹, 30% of solid tumours have mutations in Ras or Raf (ref. 2). To determine if the novel regulators isolated following RTK stimulus can also modulate ERK activation downstream of oncogenic Ras signalling (*Ras*^{V12}), we screened our validated hits in a stable S2 cell line conditionally expressing *Ras*^{V12} (ref. 23) (Fig. 2c). We found that 85 genes suppress ERK activation following *Ras*^{V12} expression (Supplementary Table 8, 9). In parallel experiments, ectopic Ras activity was induced by pre-incubation with dsRNA targeting a *Drosophila* RasGAP, *Gap1*, mimicking loss of NF1 activity in patients with neurofibromatosis, type 1 (Supplementary Table 9). These epistasis analyses suggest where, within the canonical pathway, these genes modulated ERK activation; genes that reduce ERK activation by insulin but not by *Ras*^{V12} may function upstream of Ras or in a parallel InR–ERK pathway.

Genes influencing RTK/ERK signalling output in cell culture and *in vivo* are more likely to be critical components of the regulatory network, and thus relevant to mammalian development and disease. Through a series of *in vivo* experiments testing known alleles or transgenic RNAi and complementary DNA overexpression constructs

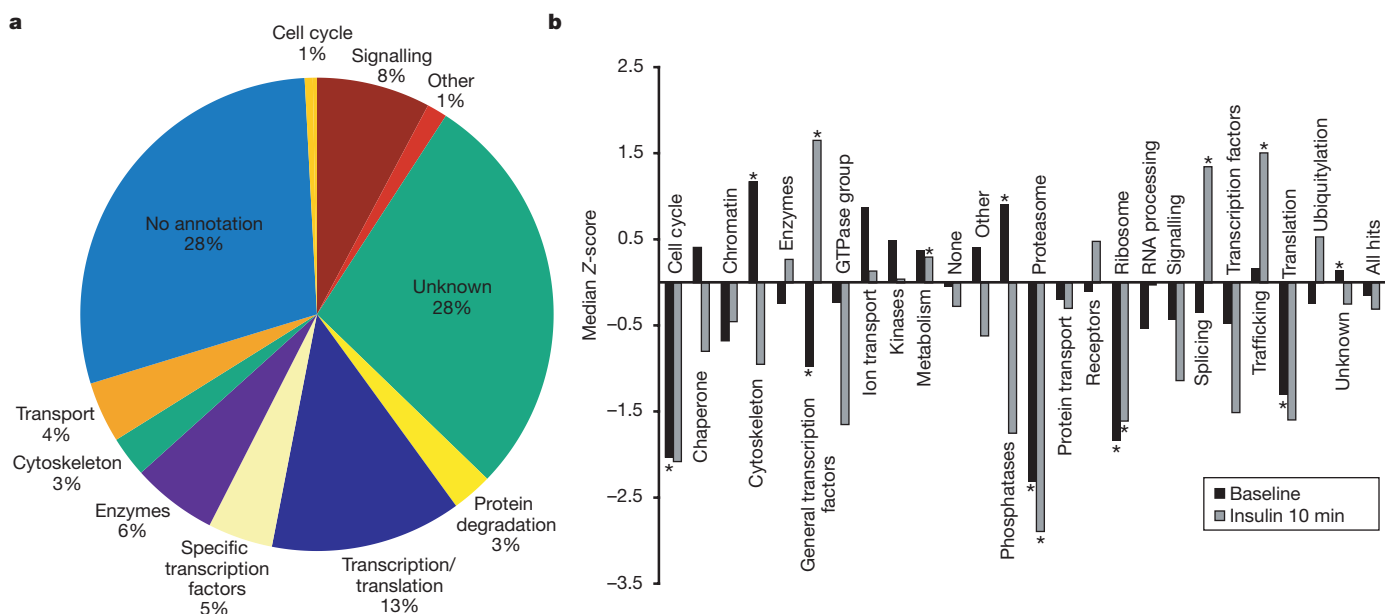


Figure 1 | Primary RTK/ERK RNAi screen results. **a**, Using gene ontology annotations, primary screen hits were grouped by general molecular function. 'No annotation' refers to predicted genes (see Methods). Isolation

of 476 predicted genes suggests that these may be valid transcripts. **b**, Median Z-scores for each functional category in the assay. Asterisk denotes $P < 0.05$.

(not shown), we focused our investigation on two uncharacterized genes, a novel positive regulator and a negative regulator of RTK/ERK signalling.

CG5169/Drosophila GCKIII (*dGCKIII*) is a member of the Ste20-like kinase group and germinal centre kinase (GCK)-III subfamily. Ste20-like kinases mediate a wide array of signalling events controlling apoptosis, proliferation and cytoskeletal organization²⁴. A *dGCKIII* knockdown reduced ERK activation in stimulated S2R+ cells (Supplementary Fig. 5a). Ras activity in the *Drosophila* imaginal disc is required for survival, proliferation and differentiation, each requiring a specific threshold of ERK activity²⁵. Wing-specific expression of a *dGCKIII* transgenic RNAi hairpin resulted in loss of tissue that was sensitive to hairpin dosage, temperature and gene dosage (Fig. 3a). This phenotype correlated with a consistent reduction in general dpERK staining (Supplementary Fig. 5d). Co-expression of

full-length *dGCKIII* cDNA, but not a kinase-dead form, rescued this phenotype (Fig. 3a), arguing that GCK-III kinase activity is required for its function.

The *dGCKIII* knockdown moderately suppressed ectopic wing veins caused by the EGFR gain-of-function (*Elp^{B1}*) background, and overexpression of *dGCKIII* in flies sensitized by heterozygosity for *Gap1* produced ectopic wing veins, suggesting that *dGCKIII* regulates ERK signalling *in vivo* (Fig. 3b and Supplementary Fig. 5e). Because cell survival (Fig. 3c), but not wing vein differentiation, seems to be the primary phenotype of reduction of *dGCKIII* gene dosage, *dGCKIII* may be required only for low-level survival ERK signalling or may function in parallel survival pathways. As other Ste20 group members have been characterized as MAP4Ks, we established by co-immunoprecipitation that *dGCKIII* could bind Raf and protein phosphatase 2A (PP2A) in S2R+ cells (Fig. 3d). As *dGCKIII*

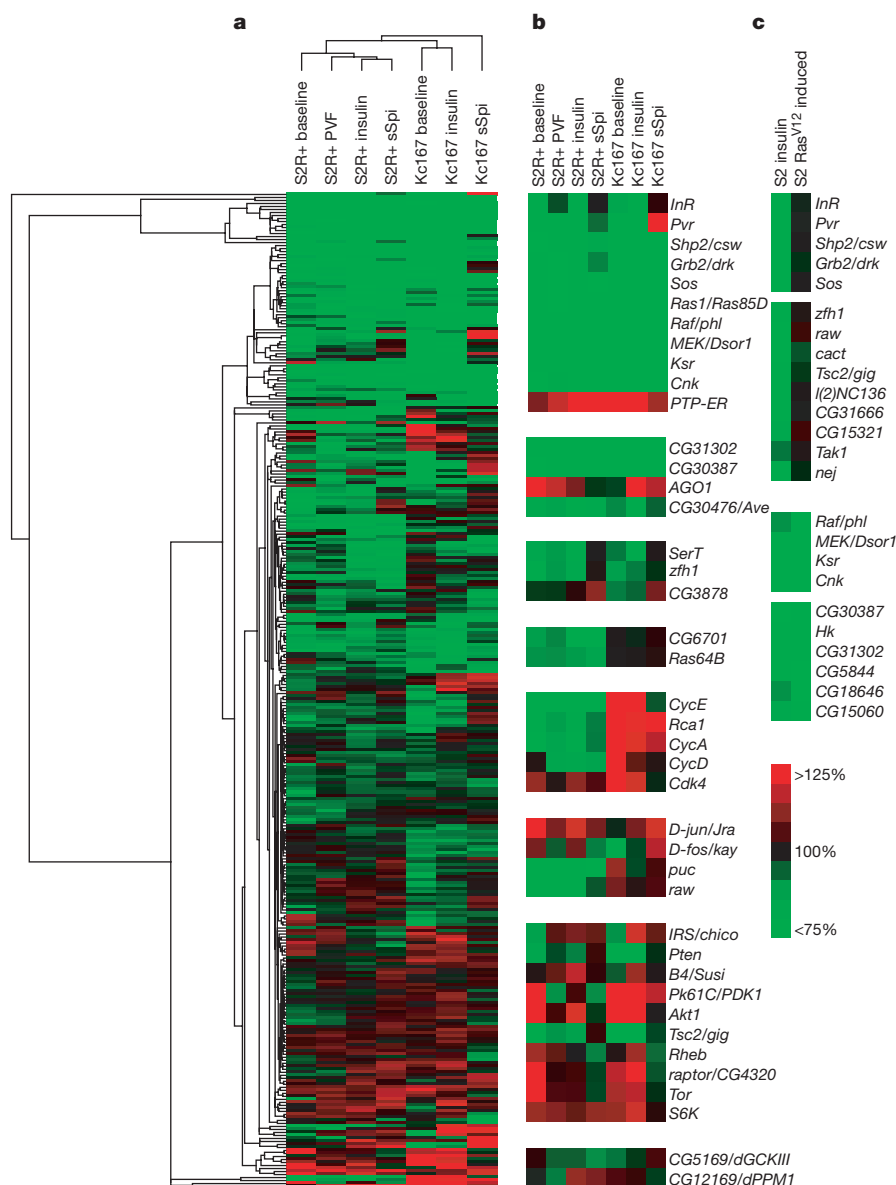


Figure 2 | Secondary screening of RTK/ERK regulators. **a**, Hierarchical clustering of hits validated in two cell lines and under multiple RTK stimuli (insulin, sSpi and PVF). Values are dpERK intensities normalized by total ERK (see Methods). Scale is per cent of a luciferase negative control; negative regulators are indicated in red and positive regulators in green. Preferential clustering of secondary screens by cell line rather than ligand stimulus suggested a large contribution of cell-line specificity to our phenotypic variability. The same clustering with associated gene names shown is

provided in Supplementary Fig. 4a. **b**, Selected groups of genes that are described in the text. **c**, Selected genes screened in an S2 cell line conditionally expressing Ras^{V12}, under insulin stimulus or following 24 h Ras^{V12} induction. dsRNAs targeting genes that suppress ERK activation following insulin stimulus, but not Ras^{V12} stimulus (as is the case for known upstream canonical components) may function upstream of Ras in the RTK/ERK pathway. Also shown are examples of genes that suppress dpERK following both RTK and Ras^{V12} stimuli.

appears to be membrane-localized (Supplementary Fig. 5g), dGCKIII may be a component of the Raf activation complex containing PP2A²⁶. Finally, siRNA-mediated knockdown of dGCKIII orthologues *MST3* and *MST4* in human prostate cancer cells reduced basal ERK1/2 activation (Supplementary Fig. 5h). As *MST4* expression is elevated in

poorly differentiated prostate cancer cells, and a kinase-dead mutant can inhibit *in vivo* tumour growth²⁷, our results suggest that GCK-III Ste20 family members may contribute to prostate cancer progression through RTK/ERK signalling.

CG1269/dPPM1 is a member of the *Drosophila* class of highly conserved PPM Ser/Thr phosphatases²⁸. Recently, a PPM family member *alphabet/CG1906* was demonstrated to regulate Ras signalling *in vivo*²⁸. A *Drosophila* PPM1 (*dPPM1*) knockdown increased dpERK in S2R+ cells after insulin activation (Supplementary Fig. 6a). Conversely, overexpression of *dPPM1* strongly suppressed ERK activation; a catalytically inactive *PPM1* failed to suppress ERK activation, but, rather, elevated dpERK levels (Fig. 4a), perhaps acting as a dominant negative.

Wing-specific overexpression of *dPPM1*, but not a phosphatase-dead form, reduced wing size and suppressed the ectopic wing vein phenotype of *Elp^{B1}* (Fig. 4b and Supplementary Fig. 6b). Overexpression also moderately suppressed the rough eye phenotype induced by Ras^{V12} under the control of the *sevenless* promoter (Fig. 4c), arguing that PPM1 reduces RTK/ERK signalling *in vivo* in a catalytically dependent manner. PPM proteins seem to be general 'activation T-loop' Thr phosphatases of substrates such as MAPKs^{29,30}. Consistent with this, we detected dPPM1 and ERK interaction by co-immunoprecipitation (Fig. 4d). The D228A mutant demonstrated increased ERK interaction, as described for a human orthologue with p38³⁰. Finally, siRNA-mediated reduction of the human orthologue PPM1 α in human prostate cancer cells increased ERK activation following EGF stimulation (Supplementary Fig. 6e). This suggests that a role of dPPM1, in particular, and PPM-family phosphatases, in general, as negative regulators of ERK activation is conserved.

In summary, a functional genomic screen for novel regulators of RTK/ERK signalling demonstrates that the precise level of ERK activation is the result of integration of other canonical pathways, a

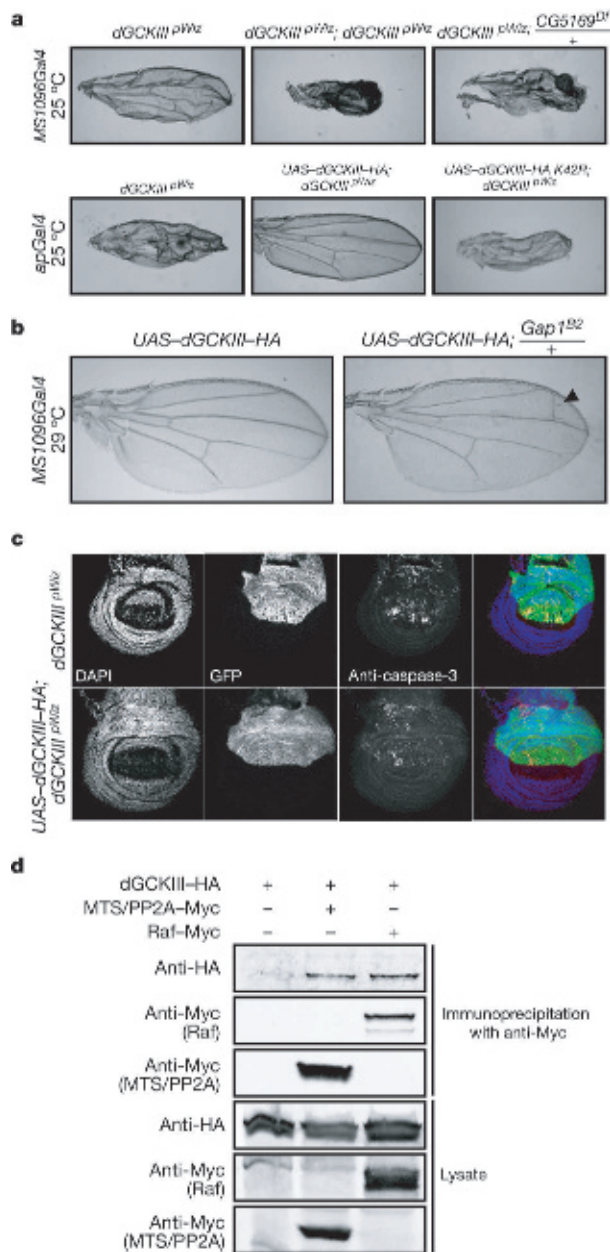


Figure 3 | Characterization of dGCKIII, a novel Ste20-like kinase. **a**, Wing-specific expression of a hairpin targeting *dGCKIII* (*dGCKIII^{Wiz}*), driven by either *MS1096Gal4* or *apGal4* as indicated. The hairpin reduced expression of HA-tagged dGCKIII in wing discs (Supplementary Fig. 5c). *dGCKIII* is ubiquitously expressed in the wing disc (Supplementary Fig. 5b). Tissue loss was increased by doubling hairpin dosage or removing one wild-type gene copy using a deficiency disrupting this gene (*CG5169^{DF}*). This phenotype could be rescued by *dGCKIII* cDNA (*UAS-dGCKIII-HA*), but not a kinase-domain-mutant (*UAS-dGCKIII-HA K42R*) *dGCKIII*, expressed to similar levels (Supplementary Fig. 5c). **b**, Overexpression of *dGCKIII-HA* in wing discs. We observed no phenotype of *Gap1^{B22}* or *dGCKIII-HA* alone (not shown). **c**, Increased cell death, detected with cleaved caspase 3, in wing discs expressing the *dGCKIII* hairpin driven by *apGal4*, marked using *UAS-mCD8-GFP*. Cell death was partially rescuable by co-expression of *dGCKIII* cDNA. No decrease in proliferation was seen (Supplementary Fig. 5f). **d**, Co-immunoprecipitation of dGCKIII by MTS/PP2A and Raf. Interaction with PP2A was predicted from yeast data sets (see Methods).

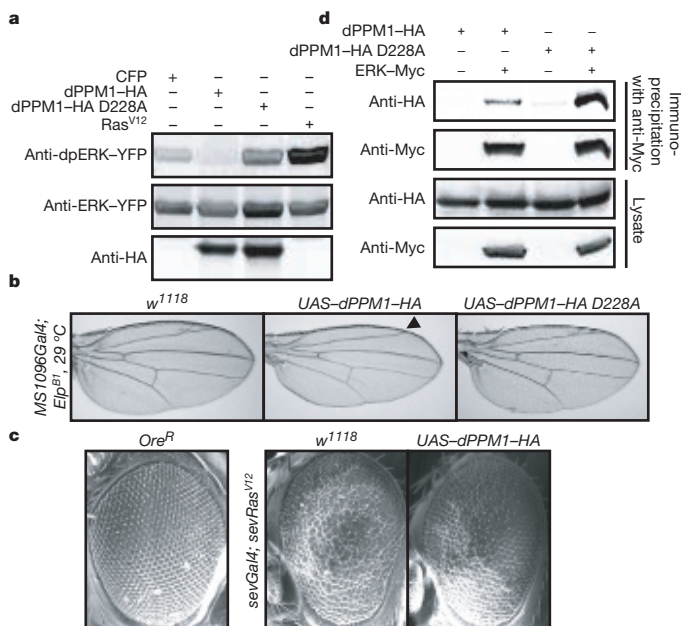


Figure 4 | Characterization of dPPM1, a novel PPM phosphatase. **a**, Overexpression of dPPM1-HA or catalytically inactive dPPM1-HA D228A with YFP-tagged ERK in S2R+ cells, following insulin stimulation. CFP and Ras^{V12} are negative and positive controls, respectively. We found an average 64% decrease in normalized dpERK signal intensity; the D228A mutant elevated dpERK levels by 31% on average. **b**, Suppression of the *Elp^{B1}* ectopic wing vein phenotype by dPPM1, but not the D228A mutant, expressed at similar levels (Supplementary Fig. 6c). **c**, Suppression of the *sevenless-Ras^{V12}* rough eye phenotype by dPPM1. *Ore^R*, wild type; *w¹¹¹⁸*, negative control. **d**, Co-immunoprecipitation of dPPM1 and dPPM1 D228A by ERK. The D228A mutant protein was consistently more efficiently precipitated with ERK than the wild-type protein.

wide range of biological processes, and extensive pathway feedback. Integrating this data with other genome-wide approaches should provide a more complete understanding of the cellular response to RTK signals. As targeted cancer therapies—that inhibit the growth of malignant cells by interfering with the causative mutant signalling pathways—begin to supplement conventional cytotoxic chemotherapy, the need to identify novel and specific drug targets in these pathways is becoming more urgent. Pathway-specific functional approaches complement phenotypic-driven screens in mammalian systems¹², both of which may lead to the identification of novel targets for drug discovery.

METHODS

The genome-wide dsRNA screening collection of the *Drosophila* RNAi Screening Center covering >95% of the *Drosophila* genome has been previously described. We observed similar kinetics of ERK activation and effects of RNAi of known pathway components in the Rolled–YFP stable S2R+ cell line compared to wild-type cells (see Supplementary Figs 1–3). Furthermore, all secondary screens were performed in wild-type cells. We adapted previous antibody-based assays⁶ for our high-throughput screen. Monoclonal dpERK antibodies (Cell Signaling) were directly conjugated to Alexa 647 dye using the Alexa Fluor 647 Protein Labeling Kit (Molecular Probes). Each time point (baseline and stimulated) was screened in duplicate.

For secondary screening, we used both total ERK and dpERK antibodies (Cell Signaling). Similar plate manipulations were used as in the primary screen, with the addition of secondary antibody incubations. Normalized dpERK values are represented as per cent of controls, and *P*-values were calculated for each gene and assay using the Student's *t*-test. Multiple hypothesis correction was applied using the False Discovery Rate (FDR) criteria ($q < 0.05$ as threshold for all assays). In addition to wild-type cells, two stable cell lines were established (S2R+ and Kc167) that conditionally express DER. Because the primary source of baseline ERK activation is through PVR, genes isolated at baseline but not under insulin stimulus may be specific effectors of PVR activity. We also screened our secondary gene list under hyper-stimulation of PVR (see Supplementary Information for secondary screen assay validation). S2 cells expressing inducible Ras^{V12} were provided by M. Therrien²³. For *Gap1* epistasis, S2R+ cells were pre-incubated with either *luciferase* (*luc*) or *Gap1* dsRNA 48 h before secondary screening.

Standard procedures were used for all other experiments. See Supplementary Information for a more detailed description of the methods.

Received 10 July; accepted 22 September 2006.

Published online 1 November 2006.

- Krause, D. S. & Van Etten, R. A. Tyrosine kinases as targets for cancer therapy. *N. Engl. J. Med.* **353**, 172–187 (2005).
- Hilger, R. A., Scheulen, M. E. & Strumberg, D. The Ras-Raf-MEK-ERK pathway in the treatment of cancer. *Onkologie* **25**, 511–518 (2002).
- DasGupta, R., Kaykas, A., Moon, R. T. & Perrimon, N. Functional genomic analysis of the Wnt–wingless signaling pathway. *Science* **308**, 826–833 (2005).
- Gabay, L., Seger, R. & Shilo, B. Z. MAP kinase *in situ* activation atlas during *Drosophila* embryogenesis. *Development* **124**, 3535–3541 (1997).
- Clemens, J. C. *et al.* Use of double-stranded RNA interference in *Drosophila* cell lines to dissect signal transduction pathways. *Proc. Natl Acad. Sci. USA* **97**, 6499–6503 (2000).
- Kiger, A. *et al.* A functional genomic analysis of cell morphology using RNA interference. *J. Biol.* **2**, 27. doi: 10.1186/1475-4924-2 (1 October 2003).
- York, R. D. *et al.* Rap1 mediates sustained MAP kinase activation induced by nerve growth factor. *Nature* **392**, 622–626 (1998).
- Luschnig, S., Krauss, J., Bohmann, K., Desjeux, I. & Nusslein-Volhard, C. The *Drosophila* SHC adaptor protein is required for signaling by a subset of receptor tyrosine kinases. *Mol. Cell* **5**, 231–241 (2000).
- Arevalo, J. C., Pereira, D. B., Yano, H., Teng, K. K. & Chao, M. V. Identification of a switch in neurotrophin signaling by selective tyrosine phosphorylation. *J. Biol. Chem.* **281**, 1001–1007 (2006).

- Johnson, S. M. *et al.* RAS is regulated by the *let-7* microRNA family. *Cell* **120**, 635–647 (2005).
- Roignant, J.-Y., Hamel, S., Janody, F. & Treisman, J. E. The novel SAM domain protein Aveugle is required for Raf activation in the *Drosophila* EGF receptor signaling pathway. *Genes Dev.* **20**, 795–806 (2006).
- Westbrook, T. F. *et al.* A genetic screen for candidate tumor suppressors identifies REST. *Cell* **121**, 837–848 (2005).
- Chang, F. *et al.* Regulation of cell cycle progression and apoptosis by the Ras/Raf/MEK/ERK pathway (Review). *Int. J. Oncol.* **22**, 469–480 (2003).
- Asha, H. *et al.* Analysis of Ras-induced overproliferation in *Drosophila* hemocytes. *Genetics* **163**, 203–215 (2003).
- Nebreda, A. R., Gannon, J. V. & Hunt, T. Newly synthesized protein(s) must associate with p34cdc2 to activate MAP kinase and MPF during progesterone-induced maturation of *Xenopus* oocytes. *EMBO J.* **14**, 5597–5607 (1995).
- Chen, X. *et al.* Cyclin D-Cdk4 and cyclin E-Cdk2 regulate the Jak/STAT signal transduction pathway in *Drosophila*. *Dev. Cell* **4**, 179–190 (2003).
- Yang, S. H., Sharrocks, A. D. & Whitmarsh, A. J. Transcriptional regulation by the MAP kinase signaling cascades. *Gene* **320**, 3–21 (2003).
- Lipsick, J. S. synMuv verité—Myb comes into focus. *Genes Dev.* **18**, 2837–2844 (2004).
- Jordan, K. C. *et al.* Genome wide analysis of transcript levels after perturbation of the EGFR pathway in the *Drosophila* ovary. *Dev. Dyn.* **232**, 709–724 (2005).
- Irie, H. Y. *et al.* Distinct roles of Akt1 and Akt2 in regulating cell migration and epithelial–mesenchymal transition. *J. Cell Biol.* **171**, 1023–1034 (2005).
- Zimmermann, S. & Moelling, K. Phosphorylation and regulation of Raf by Akt (protein kinase B). *Science* **286**, 1741–1744 (1999).
- Puig, O. & Tjian, R. Transcriptional feedback control of insulin receptor by dFOXO/FOXO1. *Genes Dev.* **19**, 2435–2446 (2005).
- Roy, F., Laberge, G., Douziech, M., Ferland-McCollough, D. & Therrien, M. KSR is a scaffold required for activation of the ERK/MAPK module. *Genes Dev.* **16**, 427–438 (2002).
- Dan, I., Watanabe, N. M. & Kusumi, A. The Ste20 group kinases as regulators of MAP kinase cascades. *Trends Cell Biol.* **11**, 220–230 (2001).
- Halfar, K., Rommel, C., Stocker, H. & Hafen, E. Ras controls growth, survival and differentiation in the *Drosophila* eye by different thresholds of MAP kinase activity. *Development* **128**, 1687–1696 (2001).
- Ory, S., Zhou, M., Conrads, T. P., Veenstra, T. D. & Morrison, D. K. Protein phosphatase 2A positively regulates Ras signaling by dephosphorylating KSR1 and Raf-1 on critical 14-3-3 binding sites. *Curr. Biol.* **13**, 1356–1364 (2003).
- Sung, V. *et al.* The Ste20 kinase MST4 plays a role in prostate cancer progression. *Cancer Res.* **63**, 3356–3363 (2003).
- Baril, C. & Therrien, M. Alphabet, a Ser/Thr phosphatase of the protein phosphatase 2C family, negatively regulates RAS/MAPK signaling in *Drosophila*. *Dev. Biol.* **294**, 232–245 (2006).
- Cheng, A., Kaldis, P. & Solomon, M. J. Dephosphorylation of human cyclin-dependent kinases by protein phosphatase type 2C α and β 2 isoforms. *J. Biol. Chem.* **275**, 34744–34749 (2000).
- Takekawa, M., Maeda, T. & Saito, H. Protein phosphatase 2C α inhibits the human stress-responsive p38 and JNK MAPK pathways. *EMBO J.* **17**, 4744–4752 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Philippakis and P. Hong for helpful discussions of statistical approaches, M. Melnick for dpERK antibody assistance, L. Kockel for pioneering the phospho-specific antibody high-throughput screen approach in our laboratory and help with *in vivo* analysis, C. Micchelli and R. Binari for assistance with genetic manipulations and interpretation, and other current and former members of the Perrimon lab for reagents and discussions. We thank L. Kockel, C. Micchelli, M. Kulkarni, B. Neel, R. Dasgupta, R. Binari and B. Mathey-Prevot for critical manuscript review. A.F. is a recipient of the Medical Scientist Training Program (MSTP) grant. N.P. is an investigator of the Howard Hughes Medical Institute.

Author Information Full datasets and dsRNA sequence information are available at the DRSC website (<http://www.flyrnai.org>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to N.P. (perrimon@receptor.med.harvard.edu).

CORRIGENDUM

doi:10.1038/nature05297

Earthquakes triggered by silent slip events on Kīlauea volcano, Hawaii

Paul Segall, Emily K. Desmarais, David Shelly, Asta Miklius & Peter Cervelli

Nature 442, 71–74 (2006)

There was a plotting error in Fig. 1 that inadvertently displays earthquakes for the incorrect time interval. The location of earthquakes during the two-day-long slow-slip event of January 2005 are shown here in the corrected Fig. 1. Because the incorrect locations were also used in the Coulomb stress-change (CSC) calculation, the error could potentially have biased our interpretation of the depth of the slow-slip event, although in fact it did not. Because nearly all of the earthquakes, both background and triggered, are landward of the

slow-slip event and at similar depths (6.5–8.5 km), the impact on the CSC calculations is negligible (Fig. 2; compare with Fig. 4 in original paper). The error does not alter our conclusion that the triggered events during the January 2005 slow-slip event were located on a subhorizontal plane at a depth of 7.5 ± 1 km. This is therefore the most likely depth of the slow-slip events. We thank Cecily J. Wolfe for pointing out the error in the original Fig. 1.

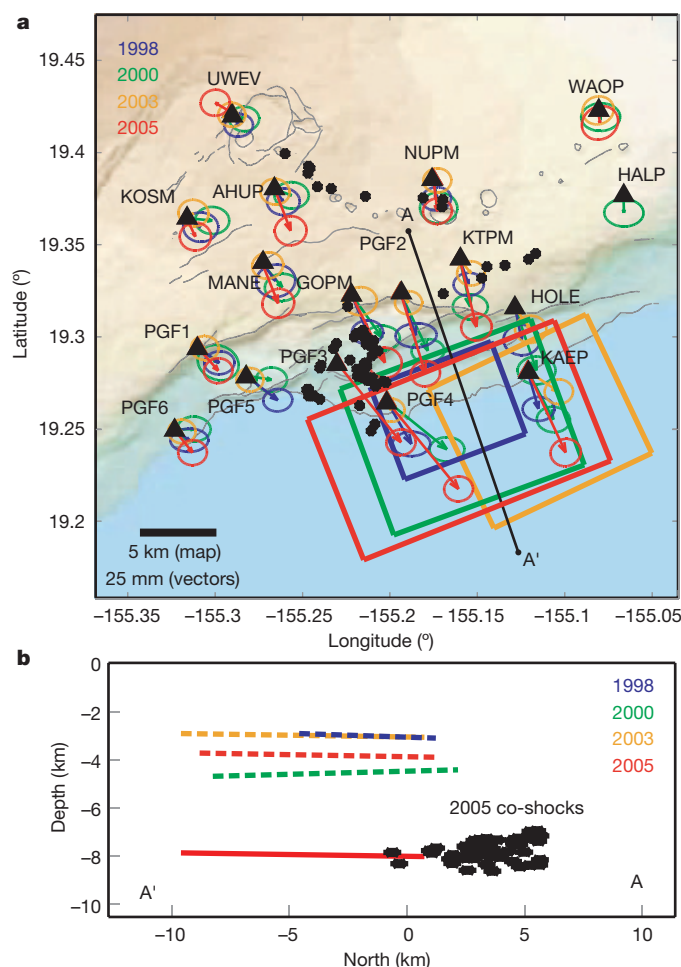


Figure 1 | Displacements and inferred slip zones for four silent slip events. **a**, Map view. Vectors indicate displacements determined as the difference between mean position before and after the event. Ellipses represent 95% confidence intervals. Rectangles show surface projections of best-fitting dislocations found by nonlinear optimization, using a simulated annealing procedure. Asterisks represent earthquakes during the 2.2 days of the slow-slip event beginning 00 h UTC on 26 January 2005. **b**, Cross-section. Dashed lines represent dislocations from inversion of Global Positioning System (GPS)-derived displacements. The solid red line indicates the 2005 event with depth constrained by seismicity. GPS station location names are given by their abbreviations.

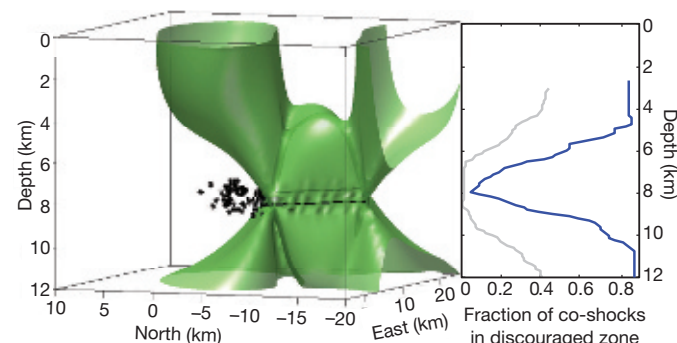


Figure 2 | Earthquake locations in relation to CSC due to fault slip. **a**, Triggered earthquakes for the 2.2-day event beginning 00 h UTC on 26 January 2005 and isosurface of constant CSC on horizontal planes. Outside the green surface the stress change encourages slip. For a dislocation surface at a depth of 4–5 km, a majority of the earthquakes fall within the zone of negative CSC. **b**, Fraction of earthquakes for which $CSC < 0$ as a function of the depth of the slow-slip event (shown in grey in the original paper, here shown in blue in the corrected figure). The minimum at a depth of 6.5–8.5 km indicates the preferred depth of the slow-slip zone.

CORRIGENDUM

doi:10.1038/nature05262

Atomic-resolution chemical analysis using a scanning transmission electron microscope

N. D. Browning, M. F. Chisholm & S. J. Pennycook

Nature 366, 143–146 (1993)

In this Letter, we reported that it was possible to undertake chemical analysis with atomic resolution in a scanning transmission electron microscope—a capability that has subsequently been demonstrated in a variety of other contexts. We now realize that, in the course of revising the manuscript in response to reviewers' comments, errors were introduced into the paper, resulting in inconsistencies in the manner in which key data were presented. In particular, the electron energy-loss spectroscopy (EELS) data central to the work (Fig. 3b) were not subject to the background subtraction processes as described in the paper, despite assurances to the contrary that we offered at the time to both the referees and the *Nature* editors. Rather, we have now concluded that only spectra 5–7 were processed in the manner described; for spectra 1–4, owing to an error, the data were reproduced from ref. 7, where a standard exponential background subtraction was used. The exponential background subtraction method is widely used in the field and is an entirely acceptable method of analysis, and therefore this error in no way affects the key scientific claim of the paper, namely that it is possible to perform atomic-resolution chemical analysis in the scanning transmission electron microscope. We sincerely regret this error and any confusion that may have resulted.

ERRATUM

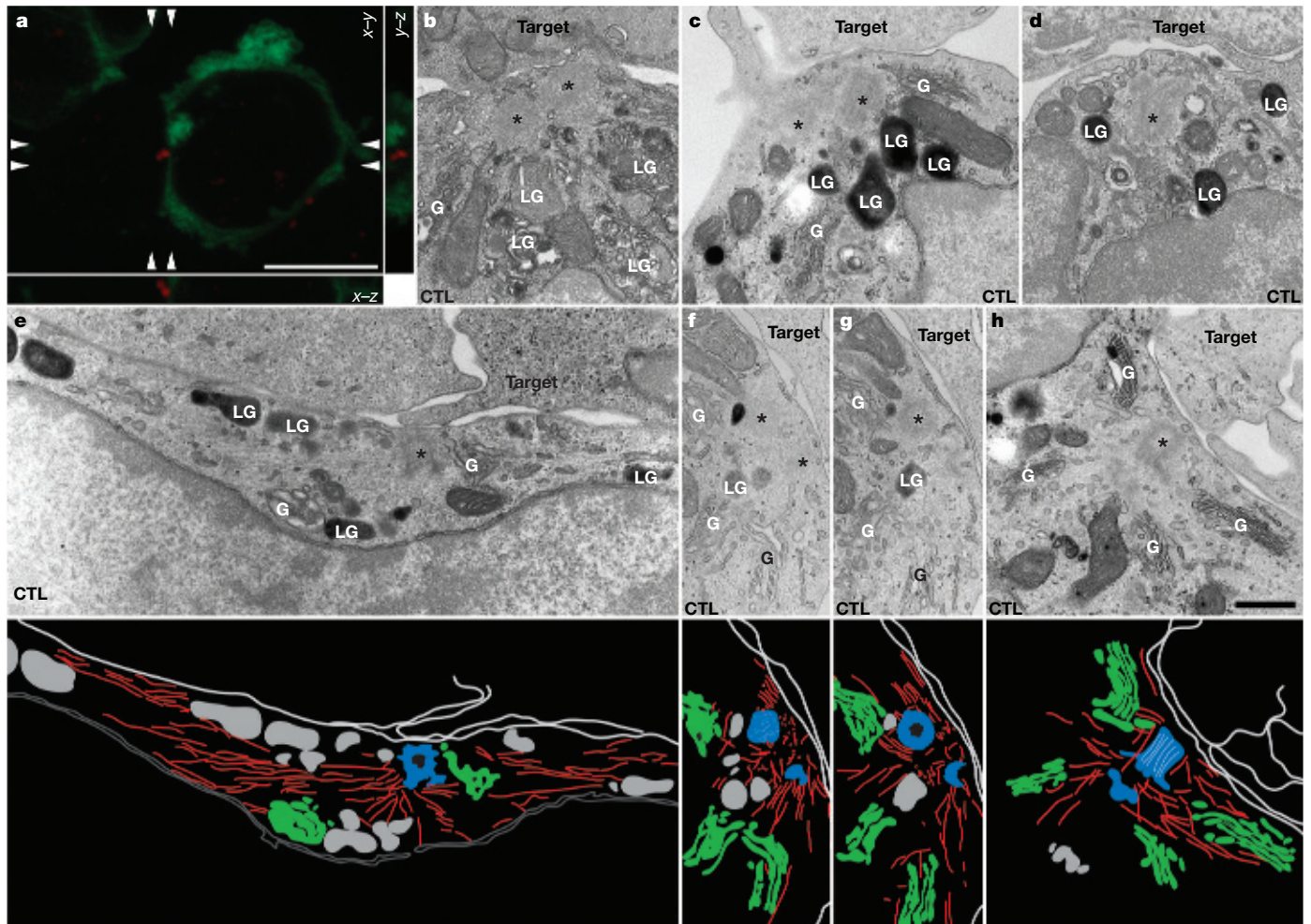
doi:10.1038/nature05299

Centrosome polarization delivers secretory granules to the immunological synapse

Jane C. Stinchcombe, Endre Majorovits, Giovanna Bossi, Stephen Fuller & Gillian M. Griffiths

Nature 443, 462–465 (2006)

In the print and PDF versions of this Letter, Figure 2 was printed incorrectly. Panels d and h were missing and part of panels c and g were removed. The online HTML version is correct. Figure 2 is reprinted correctly below.



naturejobs

**THE CAREERS
MAGAZINE FOR
SCIENTISTS**

One of the most common questions I've been asked at careers panels in the past few years is: "How can I get into industry?" The frequency is not limited by geography: I've heard this same question in the United States, France, Germany, Britain and Japan. This regularity underscores the need for young scientists to learn about the career options that exist outside academia. It also implies that they erroneously view industry as some kind of a monolithic entity, a sort of black box. This week, *Naturejobs* concludes a year-long series about industry jobs — what they are, how to prepare for them, where they can lead — with a feature that focuses on sales and marketing (see page 238).

Perhaps the single biggest lesson that can be drawn from this series is that industry offers a huge variety of opportunities. Jobs exist at many different stages of product research and development, and on both sides of the bench. Each stage requires a different set of skills, and has its own entry points. For example, pharmaceutical companies working on small-molecule drugs tend to hire synthetic inorganic chemists straight out of graduate school and then train them to make the kinds of molecules they seek. But someone in drug discovery — an earlier stage of the product pipeline — is more likely to come from a senior academic position or from within the corporate ranks.

Off-the-bench areas also exist at multiple points in the pipeline. Skills such as regulatory affairs and intellectual-property development come into play as products are being discovered, designed and developed. 'Soft skills', such as project planning and communication, can bridge the gap between the various teams in an enterprise; and scientists who have both those skills as well as a 'hard' research background can excel away from the bench.

We hope that our exploration of the functions and skills throughout the pipeline (see www.naturejobs.com) has given scientists a better understanding of their career options, and will make it easier for them to move in or out of industry, up and down the pipeline, and from one side of the bench to the other.

Paul Smaglik, *Naturejobs* editor

Publisher: Ben Crowe
Editor: Paul Smaglik
Assistant Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street, London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com
Natureevents:
Claudia Paulsen Young (+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com
France/Switzerland/Belgium:
Muriel Lestringuez (4994)

UK/Ireland/Italy/RoW:
Loredana Milanese (4944)
Nils Moeller (4953)
Scandinavia/Spain/Portugal:
Evelina Rubio-Morgan (4973)
Germany/Austria/The Netherlands:
Reya Silao (4970)
Online Job Postings:
Matthew Ward (+44 (0) 20 7014 4059)

European Satellite Office
Germany: Patrick Phelan
Tel: +49 89 54 90 57 11
Fax: +49 89 54 90 57 20
e-mail: p.phelan@nature.com

Advertising Production Manager:
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs web development: Tom Hancock
Naturejobs online production:
Catherine Alexander

US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
Chiyoda Building, 2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe
e-mail: a.watanabe@natureasia.com

Closing the deal

Sales and marketing jobs at pharmaceutical companies offer the opportunity to combine science with social skills and creative flair.

Hannah Hoag investigates the pitch.

For Georgina Smith, a career in medical sales combined her dual interests in science and business. After completing her bachelor's degree in biomedical science, Smith joined AstraZeneca as a junior medical representative. Today she drives around London visiting physicians, pharmacists, nurses and practice managers, promoting gastrointestinal and respiratory treatments.

Smith feels her science training gives her an advantage in sales. "Not everyone has that background. It allows me to talk on the same level as the physician and to understand the broader picture," she says.

Like Smith, many science graduates choose to move into the pharmaceutical industry but leave the lab work to others. Sales and marketing careers allow science graduates to interact with a variety of people in the pharmaceutical chain, from researchers and regulators to physicians and patient groups.

In the United States, the pharmaceutical industry employs more than 100,000 people to pay visits to physicians and promote medicines. There are another 100,000 or so of these sales representatives in Europe. This face-to-face contact sells billions of dollars worth of drugs annually. From 2000 to 2005 the number of sales reps continued to grow — by as much as 60% in Europe. But now companies are downsizing teams and investing in sales quality instead.

"Sales drives the bottom line for these corporations," says Tom Ruff, president and chief executive of Tom Ruff Company, a firm based in Manhattan Beach, California, that specializes in medical device and pharmaceutical sales, and sales management recruiting. "The sales representative plays a critical role."

Sell culture

The primary purpose of a pharmaceutical sales rep is to promote the company's products to customers. These are usually groups of physicians, hospitals or other healthcare providers. In these entry-level positions, the rep learns about regulatory guidelines and presenting the product to the customer. They are generally expected to make eight calls per day (see 'A day in the life of a sales rep').

"As the sales rep moves up in level from entry to senior medical representative, they become more knowledgeable about their customer market, take responsibility for pre- and post-call planning, and have more advanced analytical skills," says Michele Crocco, director of human resources, commercial operations at Roche in Nutley, New Jersey. They may become divisional sales managers, she says, coaching and evaluating other sales reps.

A career in sales requires the right sort of personality. Those who excel are outgoing people who enjoy the challenge of "making the close", says Chris Jock, vice-president and general manager for global



operations at Kelly Scientific Resources in Troy, Michigan. They are persuasive, competitive and confident self-starters. Ultimately, performance is judged on the number of prescriptions sold and contacts made, which can make pharmaceutical sales a high-pressure job.

Even entry-level sales positions frequently demand previous sales experience. "Companies want a proven track record of success in sales," says Ruff, who adds that experience selling in the healthcare industry isn't necessary. In fact, Ruff recommends fresh graduates gain a couple of years' sales experience at a Fortune 500 company known for its sales training before entering pharmaceutical sales.

But people with science backgrounds, or who are good at communicating a message and building relationships, may be able to talk their way into an entry-level position without any sales experience. Clinical experience, in particular, is rapidly becoming a valuable asset for prospective pharmaceutical sales reps.

Marketing jobs in the industry offer another option for scientists with bachelor's or graduate degrees. Well suited to people with a creative streak, these positions can be found both within drug companies and at advertising agencies and communications companies.

Marketing can take many forms, including direct

IMAGES.COM/CORBIS

mail and journal advertisements, direct-to-consumer campaigns on television, and developing sales training materials.

Rarely will a recent graduate start at an entry-level marketing position in a pharmaceutical company. Most have external marketing experience, or are hired to marketing positions from other parts of a company, including sales, because they know the company, the products and the sales force. Some large companies offer internships or graduate-trainee programmes in marketing as a way to gain experience. AstraZeneca brings in up to two dozen recent graduates a year to learn the sales and marketing ropes.

A brand associate helps the marketing team develop a communication strategy, researches and reviews brochure and website material, and observes focus groups to determine a marketing strategy's strengths and weaknesses. Brand managers generally have a few years of experience working with the brand. They develop marketing plans and promotional campaigns, analyse market research and liaise with key opinion leaders. Marketing professionals often work on products that are years away from being launched.

"You can be the first person to shape a brand, years before it is available to the public," says Simon Appleby, vice-president of human resources, global marketing, at AstraZeneca.

Those interested in pharmaceutical marketing may begin their job searches at advertising agencies and healthcare communication companies, where copywriting and medical-writing jobs are abundant. Copywriters help develop strategic messages to promote a product to a specific audience, and they may even get to name a drug. Medical writers develop sales training materials and education information for patients and physicians.



Turning pipeline products into sales: (from left) Georgina Smith, Chris Jock and Tom Ruff.

A science degree can help launch a sales or marketing career. Sales recruiters are reporting an increased demand for people with a strong scientific or clinical background — such as knowledge of anatomy, microbiology, physiology, diseases and pharmacology. And most advertising agencies and communication companies look for writers with master's-level life-sciences degrees who are able to read and interpret clinical studies.

"We have more graduates in sales and marketing with science degrees than other degrees," says Andrew Armes, head of talent development at AstraZeneca. "We attract people interested in science who want to make a difference in whatever type of role they pursue."

Letters of introduction

Although an advanced degree doesn't guarantee success in a career that is evaluated on performance, originality and creativity, it can signal a desire, and an aptitude, to be in a leadership or supervisory role. It can also bestow credibility in a physician's office, says Ruff. Those few extra letters — PharmD or PhD — can set the tone for the visit.

"Having the technical background helps refine the message," says Jock. According to *Pharmaceutical Representative* magazine's 2006 *Access Report*, physicians respond better to sales representatives who demonstrate an excellent knowledge of the condition for which the product is used, and have the ability to present clinical data in a clear and easily understood manner.

Even so, a sales or marketing career has uncertain prospects. The recent spate of mergers and acquisitions, and pressure to cut costs, could threaten job stability in the field. "A company can't sustain having 6,000 sales people, especially if you have a diminishing pipeline," says Jock. Industry analysts predict that sales forces will be cut in the future. A 2006 study by Roland Berger Strategy Consultants, *The European Pharmaceutical Industry: Delivering Sales Excellence in Turbulent Times*, found that 31% of managers surveyed expected to "dramatically reduce their sales headcount over the next two years".

But Ruff and the others feel that the number of jobs will remain constant. Instead of hiring more talent, companies are focusing on building a more efficient sales force. Jock expects this to include a shift in the role of the sales rep, with more emphasis placed on marketing in response to tighter budgets. Ultimately, securing a sales or marketing position will depend on the quality of your selling skills. If you can sell yourself, you can sell the product.

Hannah Hoag is a freelance writer based in Montreal, Canada.

A DAY IN THE LIFE OF A SALES REP

Pharmaceutical sales representatives typically begin the day by assembling a list of physicians to visit and going over their sales pitch. They make an average of eight calls each day, visiting physicians, nurse practitioners, physicians' assistants and other healthcare professionals.

Sales reps must use their time more

efficiently than ever, as physicians are pressed to see more and more patients. Historically, pharmaceutical sales reps had the luxury of 10-minute visits. Today visits are usually three to five minutes.

"They need to introduce themselves, get the message across, close and deliver the samples," says Michele Crocco (left), director of human resources, commercial operations, at Roche. "Access is a lot more difficult today, and so they have to be prepared."

In addition to selling, sales reps keep customer lists, take notes during calls, write reports for managers, attend meetings and maintain expense logs. They also organize promotional materials, drug samples and lunchtime seminars for medical specialists. The sales rep's day rarely ends at 5 p.m. There may be an educational programme for physicians to host over dinner, or administrative work to be completed. **H.H.**



MOVERS

Georgina Mace, director, Centre for Population Biology, Imperial College London, UK



2000–06: Director of science, Institute of Zoology, Zoological Society of London, London, UK

1999–2000: Senior research fellow, Institute of Zoology

1994–99 NERC Advanced Research Fellow, Institute of Zoology

1991–94: Pew Scholar in Conservation & Environment, Institute of Zoology

Georgina Mace was among those attracted to John Maynard Smith's pioneering work in evolutionary ecology at the University of Sussex in the 1970s. As one of the first generation of conservation biologists, Mace has not only contributed to the scientific development of the field, but actively translated science into effective policy.

Mace's PhD thesis focused on the evolutionary ecology of small mammals, one of the first studies using comparative methods to test ecological and evolutionary hypotheses. She then went to the Smithsonian Institution in Washington DC to gain more qualitative insights from John Eisenberg, a mammalian expert at the US National Zoological Park. While there, Mace managed the effects of inbreeding in zoo populations, a task well suited to her background in evolutionary theory.

Mace moved to the University of Newcastle upon Tyne as a research fellow in comparative biology, but soon decided to work with zoo populations again. In the 1980s, she and her colleagues helped further the field by developing theories related to population viability, an area for which zoo populations were an important case study. "It was exciting to make quantitative scientific contributions to conservation," she says.

Mace continued her work at the boundary of pure and applied science, as scientific adviser and conservation coordinator of the Federation of Zoological Gardens and at the Institute of Zoology in London. There, she helped to revise the rules for the World Conservation Union's list of threatened species. In 2000 she became science director of the institute, and reorganized the disciplinary groups into thematic areas — such as wildlife epidemiology — where breakthroughs in conservation biology were taking place.

On 1 November, she became director of the Natural Environment Research Council's Centre for Population Biology at Imperial College London. Evolutionary theorist Paul Harvey, who is currently secretary of the board of the Zoological Society of London, says that Mace's professional credentials and ability to influence policy will be major assets as she takes on her first major challenge. She will be faced with creating a strategic research plan to ensure the centre's government funding and to align its priorities with national and international science needs.

Mace also plans to use Imperial's focus on environmental issues such as climate change to promote cross-disciplinary work involving biodiversity and population biology. ■
Virginia Gewin

SCIENTISTS & SOCIETIES

The view from Russia

As a scientist working and studying in Russia for the past 25 years, I've noted many changes as the country begins to adapt to new realities. Superficially the situation simply stagnates, but a more attentive view shows that it may be changing in some important ways.

Instead of monolithic control of science, we have a highly visible conflict between the Ministry of Science and the Russian Academy of Sciences. This is a good thing. As both try to appeal to the public, they serve as each other's watchdogs, and so try to avoid obvious transgressions.

On the surface at least, the ministry looks more merit-based. Its strategy — as reflected in ministry documents, public statements and interviews — includes more competitive grants (as opposed to direct budgeting), peer review, international experts, and a merit-oriented hiring and salary system. But my observations and discussions with colleagues suggest that many of its programmes are still infested with kickbacks and closed-door decisions.

The academy's governing body, the presidium, while paying lip service to peer review, is, in fact, opposed to the grant system, and relies on 'scientific programmes', many of which create a conflict of interest: decisions are made by people who directly benefit from them. Clashes between the

ministry and the presidium reached top levels over 'bonuses for productivity'. In May, the ministry decided to establish a system of merit-oriented bonuses based on impact factors and citation indices, talks at international conferences and other easily measurable criteria.

The academy's top management has attacked this project, suggesting that all decisions should be left to the discretion of institute directors. Admittedly, the ministry scheme is a temporary fix in the absence of a real peer-review system. Nevertheless, I think, it is a step in the right direction.

The role of international bodies has also changed. Few Russian scientists now depend on international grants, as in the 1990s. Still, as international agencies are usually free from local influences, grants such as Howard Hughes Medical Institute international scholar awards provide a much-needed example of fair, transparent competition.

Changes are slow and painful, but given our country's history, it would be naïve to expect otherwise. ■

Mikhail Gelfand is head of the Bioinformatics Center of the Institute for Information Transmission Problems at the Russian Academy of Sciences and a Howard Hughes International Research Scholar.

GRADUATE JOURNAL

The path to a PhD

As I work towards my PhD, I can't help but look back on how I got here. When I was 11 or so, I saw the film *Lorenzo's Oil*. The excitement of finding a cure for a disease captured my imagination and made me want to be a doctor. Learning of the cell structure, DNA and genes at school, I was enthralled. At the age of 17, after visiting hospitals and starting to apply to medical school, I decided I wanted to investigate the causes of genetic diseases rather than treat the symptoms. A degree in biology followed.

While at university, I discovered evolution on a big scale: the subtleties and power of it, unrealized at school, astonished me. Developmental genetics also interested me — how a single cell develops into a complete organism. Evo-devo was the area that would allow me to combine the two.

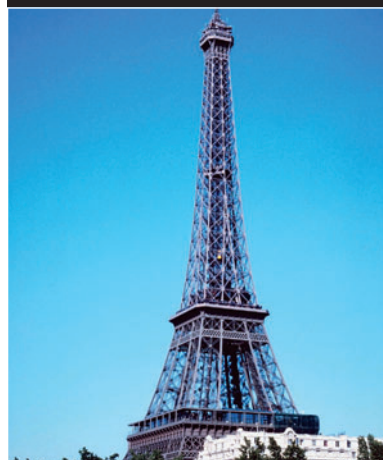
Plants seemed the best model system to ask how genetics shapes the development of different organisms. Not only are there plenty of advanced research techniques available, but plants are pleasant to work with. Unlike mice, worms or flies, leaves of different shapes and sizes surround me every day as I walk to university. How could I not be fascinated by their evolution? Sometimes I wonder what it'd be like to do more medical work. But give me a greenhouse full of ferns over an incubator of mammalian cell cultures any day. ■

Mhairi Dupré is a first-year PhD student in evolutionary developmental biology at the University of Oxford, UK.

THE HIGHLIGHT ON FRANCOPHONE COUNTRIES



in partnership with



Naturejobs

Project Manager:
Amélie Péquignot

T: +44 (0) 20 7843 4974
F: +44 (0) 20 7843 4996
E: a.pequignot@nature.com

Naturejobs

Europe:
Muriel Lestringuez

T: +44 (0) 20 7843 4994
F: +44 (0) 20 7843 4996
E: m.lestringuez@nature.com

Naturejobs/Natureevents

Canada:
Helen Wu

T: +1 212 726 9347
F: +1 212 696 9482
E: h.wu@natureny.com

Natureevents

Europe:
Claudia Paulsen Young

T: +44 (0) 20 7014 4015
F: +44 (0) 20 7843 4996
E: c.paulsenyoung@nature.com

naturejobs

natureevents

nature publishing group **npg**



The core research and development aims of the AP-HP are to form industrial partnerships, to encourage technology transfer and to support the creation of innovative companies. Industrial partnerships are largely developed through the numerous clinical trials that AP-HP conducts, but also through trials initiated by industry, thanks to the recent establishment of a “unique desk” in the Department of Clinical Research and Development (DRCD) for industry-initiated clinical trials.

The AP-HP has primarily provided its support for technology transfer and industrial

partnerships to high-quality teams developing translational programmes – from upstream research to the implementation of diagnostic and therapeutic strategies, including all the intermediate links that are indispensable to clinical research.

The proximity of the basic research and the clinical investigation teams (Centre d'Investigation Clinique AP-HP – Inserm), together with a coherent internal policy and the key strengths in research and innovation of every hospital that is part of the AP-HP network, is a major factor benefiting biomedical research.

The AP-HP not only plays a key role in fostering translational research – its activities also represent a major economic advantage for Paris and its region. The industrial partnership platforms of the AP-HP are part of the tools that form the basis for this dynamic economic development, and generate considerable employment.

Major medical and biotechnological innovations are discovered by the AP-HP teams. At present, these teams have generated an active portfolio of more than 250 patents, providing the foundation for technology transfer to external industrial partners and the creation of biotech companies to develop and exploit them.

On the industrial side, 63% of the more than one thousand clinical trials conducted each year in France by the pharmaceutical industry have at least one participating centre that involves a medical team from AP-HP.

Many bridges have been put in place by AP-HP through its proactive policy. The mission of AP-HP, reinforced by its partnership with Paris Biotech Santé, the Cochin incubator, represents a coherent and structured unit promoting innovation and the support of projects essential to clinical research.

As a key partner with public and private organizations, and one whose goals are well aligned with the development of the Medicen Sante pole of competitiveness, the AP-HP is delighted to be at the heart of the numerous ambitious projects that are transforming the Paris Ile de France region into an internationally recognized prime location for translational research and development.

Testimonial

“The proximity of hospital-university expertise and evaluation to the Industrial Partnerships Platform and the incubators is a major asset for the conduct of clinical trials of drugs and medical devices. The results of clinical trials conducted by the AP-HP (more than 470 clinical studies in active file) allow the registration of drugs with regulatory agencies in France and abroad.” Nicolas Best, General Manager of the Clinical Research and Development Department.



Nicolas Best, General Manager of the Clinical Research and Development Department.

AP-HP in Paris and suburbs:

29% of the care offer in the area
23 550 hospital beds, 1 440 day-hospital beds,
820 home care beds

Medical activities:

1 051 000 admissions/year
Stays lasting 7 days on average
5,4 million ambulatory visits
An admission every 29 sec in emergency units
32 084 births, 1 033 transplants

Personnel:

82 430 full time employees
12 330 physicians in 750 departments
57 800 auxiliary nursing staff
9 630 administrative staff
12 360 technical and operative staff

Formation:

37 training centres, including 19 nursing schools

Medical teaching, in partnership with:

11 medical schools
2 odontology and 2 pharmacy schools

Medical research, in partnership with:

100 INSERM research units (National Institute of Health and Medical Research)
30 CNRS units (National Centre of Scientific Research)
7 Clinical Investigation Centres (CIC)
8 Clinical Research Unit (URC) for methodology and statistic analyses
5 CEA unit (National Atomic Research Centre)
1 Imaging Research Centre in collaboration with INRA (National Agronomy Research Institute)

As a public health institution, biomedical and clinical research is one of the main objectives of AP-HP. This activity is driven by the Department of Clinical Research and Development (DRCD).

In collaboration with the most famous French research institutions, (INSERM, CNRS, CEA, Pasteur Institute) and medical Universities in Paris city, AP-HP teams develop high quality research activities. Within the institution, DRCD main objective is to develop institutional clinical research in order to actively contribute to innovative therapeutic and diagnostic strategies.

Expertise and know-how provided by AP-HP teams cover diagnosis and innovative therapies – including new drugs, gene and cell therapies, orphan diseases management, drug discovery, therapeutic targets characterization, highly qualified Biobanks building up, medical devices, physiopathology and clinical studies, epidemiology and human genetics.

AP-HP is the largest clinical trial center in Europe

DRCD, since its creation in 1992, has been promoting 2 000 biomedical research projects. At the steady state, DRCD is ensuring promotion of 500 projects/year, representing recruitment, monitoring data management and analysis of 12 300 people. This promotion is covering financial and scientific sponsoring of clinical studies, implementation and follow up of regulatory, financial and quality insurance processes, AP-HP research policies definition and implementation, partnerships management with French research institutions and industrial partners, and, finally, European projects follow up. For this latter purpose, DRCD has created the European Affairs Group that has already supported and managed about fifty projects.

The clinical research activity is covering all therapeutic areas – infectious, oncology, neurology, inflammatory, cardiology, paediatrics. Moreover, DRCD is managing multicenter clinical studies promoted by industrial partners. This activity is



concerning drug discovery phase I to IV, medical devices and vaccine clinical studies, 90% being at the international level.

www.drcc.aphp.fr

1. AP-HP intellectual property assets management

AP-HP holds a pipeline of 246 active patents (as full or co-owner with other private or public research institutes) in various fields: health care delivery, medical device, gene and cell therapies, drug discovery, genetics and diagnostics. AP-HP is also owning about a hundred technologies and know-how, 40 software and databases (including cellular and DNA databases), deposit marks and models (medical devices). OLTT (office of licensing and technology transfer) is examining 50 patent filing projects/year.

2. Technology Transfer and Licensing

OLTT is continuously prospecting for industrial partners, and building partnership through technology transfer and license agreements, research contracts or Material Transfer Agreement (MTA). OLTT is also promoting industrial partnerships for the new drugs and galenic innovations developed by the EP-HP (the AP-HP's pharmaceutical company) and has obtained marketing authorizations applications (AMM) from French Medicine Agency (AFSSAPS) and from EMEA. EP-HP has a portfolio of compounds which can be approved in the field of rare diseases as orphan drugs.

By the way, from its medical staff and specialities to its technical and administrative board, AP-HP represents a strong and expert partner for any kind of collaboration within biomedical and clinical research fields and offers a very high potential for business developments related to medical, biomedical and clinical activities.

Since its creation in 1993, the Office of Licensing and Technology Transfer (OLTT) is increasing the value of innovations introduced by AP-HP employees, through three main activities.

3. Industrial partnership platform

To promote the innovation and technology transfer, AP-HP provides laboratories and office space for companies. The main objective here is to increase joint research and development projects between life sciences companies and hospital's and/or Universities' teams: it must be a real public-private scientific collaboration. Within the framework of this collaboration, companies can access to private and common laboratories, biomedical research structures and patient and administrative support.

The first industrial partnership platform currently in function is based at Saint-Louis hospital for companies with R&D projects in the field of hematology-oncology-transplantation. The future industrial partnership platforms and their thematics:

- neurological and motor handicap (Raymond Poincaré Hospital)
- lengthening of human life (Charles Foix Hospital)
- institute of the brain and marrow (Pitié-Salpêtrière Hospital).

www.valorisation.aphp.fr

W90915R

McGILL UNIVERSITY

School of Computer Science

Assistant Professor

The School of Computer Science at McGill University wishes to invite applications for a tenure-track position at the assistant professor level, to begin July 1, 2007. Applications for more senior positions are also welcome. We are searching for a computational biologist who is interested in one of the following areas: algorithms in bioinformatics, machine learning/statistical inference in bioinformatics, dynamical models in bioinformatics, and bioinformatics database and system design.

Bioinformatics at McGill University has greatly expanded over the past six years and candidates would benefit from the world class medical school and biomedical research programs. The McGill Centre for Bioinformatics is comprised of approximately 16 members who are involved in large scale projects that include (but are not limited to) areas such as genomics, epigenomics, functional genomics, proteomics, structural biology, chemi-informatics, clinical informatics, modelling in physiology, and evolution.

Complete hardcopy or pdf format applications, including a curriculum vitae, a list of publications with copies of one or two sample reprints, a research statement as well as a teaching statement, and the names and e-mail addresses of three references should be sent to biosearch@cs.mcgill.ca or mailed to the address below:

Bioinformatics Search Committee
School of Computer Science
3480 University Street, Room 318
Montreal, QC H3A 2A7
Canada

The review of applications will begin on January 9, 2007, and applications should be submitted by that date; however, applications will continue to be accepted until March 1, 2007, or until the position is filled. McGill is the top-rated research university in Canada. It is located in the heart of vibrant, multicultural Montreal.

For further information about our department, see: <http://www.cs.mcgill.ca> and for information on the McGill Centre for Bioinformatics, see: <http://www.mcb.mcgill.ca>

We encourage all qualified candidates to apply; however, Canadians and permanent residents of Canada will be given priority. McGill University is committed to equity in employment.

NW90606R

IN 2007

CNRS IS RECRUITING

MORE THAN 400

TENURED RESEARCHERS

IN ALL SCIENTIFIC FIELDS*

*MATHEMATICS • PHYSICS • NUCLEAR AND HIGH-ENERGY PHYSICS
 • CHEMISTRY • ENGINEERING SCIENCES • COMMUNICATION AND
 INFORMATION TECHNOLOGY AND SCIENCES • EARTH SCIENCES AND
 ASTRONOMY • ENVIRONMENTAL SCIENCES • LIFE SCIENCES
 • HUMANITIES AND SOCIAL SCIENCES

This recruitment campaign is open to junior and senior researchers from all over the world. One of the major objectives of this campaign is to encourage international scientists to apply to CNRS.

CNRS researchers work in an enriching scientific environment:

- numerous large-scale facilities
- highly-skilled technical support
- multiple networks throughout Europe and across disciplines
- access to university research and teaching
- lab-to-lab and international mobility

At CNRS, the long-term vision of excellence in basic research provides a solid foundation for cutting-edge technological research. Successful candidates to the CNRS benefit from the dynamics, stability and stimulation of belonging to a major research organization.



As of December 2007,
online application forms
and further information
available from

www.cnrs.fr

W90795R

TENURE-TRACK POSITION IN

BIOLOGICAL PHYSICS

UNIVERSITY OF OTTAWA

The Department of Physics invites applications for a tenure-track position in experimental or theoretical biological physics. The candidate will have a Ph.D. or equivalent, an outstanding research record and a strong interest in teaching. The appointment will normally be at the Assistant Professor level, but applications for higher ranks will also be considered. The candidate will pursue a vigorous research program that will add to the Department's strengths in areas such as biological modeling and computation, neurophysics, computational biology, cellular interactions, genomics, proteomics, molecular biophysics and biophotonics. The candidate will also be part of a large contingent of researchers studying biological dynamics and computation within the Faculty of Science, the Faculty of Medicine, and its affiliated Ottawa Health Research Institutes and Ottawa Institute of Systems Biology. Computational work is further enabled by the High Performance Computing Virtual Laboratory. As Canada's National Capital, Ottawa is a vibrant and attractive city, well served for national and international travel. It has numerous cultural amenities and offers easy access to several summer and winter outdoor activities.

According to government policy, all qualified candidates are invited to apply. However, preference will be given to Canadian citizens and permanent residents. The University of Ottawa is an equal opportunity employer. We strongly encourage applications from women, Aboriginal peoples, persons with disabilities and members of visible minorities. As the University of Ottawa is a bilingual institution, bilingualism is an asset.

Applicants are requested to send a curriculum vitae, the names of at least three referees, and a statement of research interests to: Search Committee (c/o Dr. André Longtin), Department of Physics, University of Ottawa, 150 Louis Pasteur, Ottawa, Ont. Canada K1N 6N5. Applications will be reviewed starting in December 2006 and until the position is filled.

NW90626R



Faculty Position in Nanotechnology

at Ecole Polytechnique Fédérale de Lausanne (EPFL)

The School of Engineering at EPFL seeks a tenure track assistant professor in the broad field of **Nanotechnology**. Exceptional candidates at the Associate or Full professor levels may also be considered. EPFL has excellent fabrication facilities and a rich research environment in various areas of nanotechnology. The present search is for a person with a system engineering or circuits view of nanotechnology and an application focus in his/her research program. Specific areas of interest include the design of electronic circuits at the nanometer scale, nanostructured sensors, nanophotonics, nanostructured materials, NEMS, and applications of nanotechnology in biology.

The successful candidate is expected to have a deep understanding of the fabrication technology and experimental procedures as well as of modelling, design tools and methodology aspects and participate in undergraduate and graduate teaching. EPFL offers internationally competitive salaries, start-up resources and benefits.

Applications should include a curriculum with a list of publications, a concise statement of research and teaching interests, and the names and addresses (including e-mail) of at least five referees. Applications should be uploaded at <http://nano-recr.epfl.ch> by **February 15th, 2007**.

Inquiries may be sent to
Professor Giovanni De Micheli,
Chair of the Search Committee,
INF 341
Station 14
CH-1015 Lausanne, Switzerland
 E-mail: recruiting.nano@epfl.ch

For additional information on the EPFL please consult: <http://www.epfl.ch>,
<http://sti.epfl.ch>, <http://cmi.epfl.ch>.

The EPFL is an equal opportunity employer.

W90733R

WE HIGHLIGHT SCIENCE

The European Synchrotron Radiation Facility (ESRF) is Europe's most powerful light source. It is located in the unique scientific environment of Grenoble, in the heart of the French Alps.

The ESRF offers you an exciting opportunity to work with international teams on a wide range of scientific subjects using synchrotron light: Physics and Chemistry, Life Sciences and Medicine, Earth and Environmental Sciences, Surface and Materials Sciences.



Scientists
 Post-doctoral fellows
 PhD students
 Engineers
 Technicians
 Administrative staff

Have a look at our vacancies at: **www.esrf.fr**
 Contact us at: **recruitment@esrf.fr**

W85755R



UNIVERSITÉ
LAVAL

Première université francophone en Amérique, l'Université Laval est l'une des plus importantes universités du Canada. Activement engagée dans son milieu, elle offre un environnement de formation et de recherche de premier plan au cœur de Québec, ville du patrimoine mondial de l'UNESCO.

Professeure, professeur GÉNIE CHIMIQUE

FACULTÉ DES SCIENCES ET DE GÉNIE

Date de clôture du concours : le 15 décembre 2006

Date d'entrée en fonction : le 1^{er} mai 2007

www.gch.ulaval.ca/poste.html

En vertu de son programme d'accès à l'égalité, l'Université Laval entend consacrer la moitié de ses postes vacants à l'engagement de femmes. En accord avec les engagements du ministère de Citoyenneté et Immigration du Canada, cette offre est destinée en priorité aux citoyennes et aux citoyens et aux résidentes et aux résidents permanents du Canada.

www.ulaval.ca

NW90607R



Tenure-Track Faculty Position in Experimental Biophysics

Department of Physics, McGill University

We are currently seeking applications for a tenure-track faculty position at the rank of Assistant Professor in the area of Experimental Biophysics, beginning as early as September 2007.

McGill University's Department of Physics offers a rich research environment that includes the availability of advanced probes and sensors, fast imaging and correlation spectroscopies, and the access to atomic manipulation and microfabrication facilities. We are looking for an individual whose vigorous research program will enhance and complement this effort, and we value strong interdisciplinary interactions within and outside our Department. The successful candidate will be a strong teacher and an outstanding researcher. Preferably by mail, and as soon as possible, interested candidates should submit a curriculum vitae, a statement of research goals and plans, a statement of teaching interests and philosophy, and arrange for at least three letters of reference to be sent directly to

Professor Charles Gale, Chair
Department of Physics, McGill University
3600 University St., Montreal, Quebec, Canada H3A 2T8

Review of applications will begin January 3rd, 2007; applications will be accepted and reviewed until the position is filled.

The successful candidate will be supported by a generous start-up package and could be nominated for a Canada Research Chair.

All qualified candidates are encouraged to apply; however Canadian citizens and permanent residents of Canada will be given priority.

McGill University is committed to equity in employment.

NW90870R

Ateliers de formation Inserm
101 rue de Tolbiac
75654 Paris Cedex 13 France

Tel: 33 (0) 144.23.62.03
Fax: 33 (0) 144.23.62.93
ateliers@tolbiac.inserm.fr

Inserm

French Institute
of Health and Medical Research

Atelier de formation n°/Workshop 174
APPLICATIONS OF MICROT TECHNOLOGIES TO CELL
BIOLOGY AND MEDICAL RESEARCH

Organizers: Franz Bruckert (CEA, Grenoble), Bertrand Fourcade (CEA, Grenoble)

REGISTRATION DEADLINE: JANUARY 15, 2007

PHASE I: CRITICAL ASSESSMENT: MARCH 15-16, 2007

• LA LONDE-LES-MAURES (TOULON)

Audience

Physicians, pharmacists, researchers, PhD. students, postdocs, engineers and technicians with an interest in tissue engineering, medical prostheses, in vitro diagnosis and/or engaged in nanobiotechnology projects.

Lectures will be given in English.

Maximum number of participants: 80.

Programme

The programme will cover a broad range of experimental techniques, with emphasis on biological and medical applications. More specifically, the workshop will address the following points: microchambers and microfluidics devices, surface patterning at the micron scale, excitable surfaces, miniature electro-mechanical devices, novel methods for the detection of molecules at the immediate vicinity of surfaces.

With the participation of

Peter Fromherz (Martinsried, Germany), Alexandra Fuchs (Grenoble, France), Piotr Garstecki (Warsaw, Poland), Noo-Li Jeon (Irvine, USA), Ali Khademhosseini (Boston, USA), Benoît Ladoux (Paris, France), Thomas Kissel (Marburg, Germany), Thierry Livache (Grenoble, France), Patrice Marche (Grenoble, France), Buddy Ratner (Washington, USA), Hervé Rignault (Marseille, France), Thomas Stieglitz (Freiburg, Germany), Jean-Claude Voegel (Strasbourg, France), Janos Vörös (Zurich, Switzerland).

PHASE II: PRACTICAL COURSE: APRIL 23-26, 2007 • GRENOBLE

Programme

Several practicals will be proposed, illustrating applications of the methods and approaches presented in phase I. Topics are: surface patterning of extracellular matrix proteins, polyelectrolyte multilayer coating, lab-on-chip devices, surface plasmon resonance imaging techniques.

Selection

24 trainees will be selected among the participants of phase I.

W90024E



Institut de recherche
pour le développement

L'Institut de Recherche pour le Développement (IRD)

Post offer:

**Director of the Centre of Research and Surveillance on Emerging
Diseases in the Indian Ocean**

New Joint Scientific Research Venture based in Saint Denis, Réunion

Duties and functions:

- coordination of scientific research and surveillance activities on emerging diseases in the Indian Ocean region
- stimulation and setting-up of new research programmes in these fields
- management of administrative and financial aspects of the centre
- organization and leadership of a partnership network, at local and regional level within the Indian Ocean Commission and also at international scale

The post requires high-level education and training and substantial experience in the field of research in epidemiology, infectious disease prevention and treatment and other health-related disciplines. The successful candidate will have a good knowledge of multidisciplinary research on emerging diseases.

You must have gained international recognition for your research activities, organization and leadership and have considerable team management experience.

You must be bilingual French/English.

Last date for applications: 1 December 2006.

Call for applications in detail on www.ird.fr

Post location: Saint Denis, Réunion.

Type of contract: Secondment, guest research director contract, or contract and isolated post allowance.

Qualifications: Doctorate and HDR (Director of research diploma) or equivalent level

CV, covering letter and written account of publications and work conducted to be sent to:

• candidat@paris.ird.fr,
reference: poste Directeur Réunion
or

• Directeur des Personnels

Institut de Recherche pour le Développement
213 rue La Fayette, 75480 Paris cedex 10, France

W90451R



INSTITUT DE GENETIQUE HUMAINE

Group Leader Positions
Institute of Human Genetics, CNRS
Montpellier, France

The Institute of Human Genetics is one of the largest CNRS Research Centre in Molecular and Cell Biology, with particular emphasis in epigenetic regulation and human genetics. It is organized in three scientific departments: Genome Dynamics and Chromatin, Development and Differentiation, Human Genetics. Detailed information on the Institute research activities, infrastructure and facilities are available in our web site: <http://www.igh.cnrs.fr>

The Institute invites applications for either established teams or new emerging teams. Applications concerning regulation of nuclear organization and chromatin structure for gene expression or genome metabolism, RNA silencing, differentiation of stem cells are encouraged. Projects aimed to analyse human genetic diseases at the molecular level will also receive particular attention. The main criteria for the selection of the candidate will be the scientific excellence.

The successful applicants will have high quality publications, and will be expected to establish a funded research program. A laboratory space will be offered as well as full access to technical platforms, common spaces and equipment of the Institute.

Applicants should submit five to ten pages stating their background and research project, a full CV and contact details for three referees. After a first selection, selected candidatures will be invited to present their work and projects at the Institute. The decisions will be communicated by 30 September 2007, at the latest. National as well as foreign candidates are encouraged. Applications should be sent to Marcel MECHALI or Alain BUCHETON, Institute of Human Genetics, 141 Rue de la Cardonille, 34396 Montpellier cedex, France. Tel: 33-(0)499 619 917, Fax: 33-(0)499 619 920, E-mail: mechali@igh.cnrs.fr and Alain.Bucheton@igh.cnrs.fr

W90875R



**PASTEUR
FOUNDATION**

**POSTDOCTORAL
FELLOWSHIPS**

Institut Pasteur, Paris, France

Founded in 1887 by Louis Pasteur and located in the heart of Paris, the Institut Pasteur is a world-renowned private research organization. The Pasteur Foundation is seeking outstanding Fellowship Applicants. Candidates may apply to any laboratory within 10 departments: Cell Biology and Infection; Developmental Biology; Genomes & Genetics; Immunology; Infection and Epidemiology; Microbiology; Neuroscience; Parasitology and Mycology; Structural Biology and Chemistry; and Virology. See website for details.

Fellowship package is \$60,000 per year for three years. U.S. citizenship required; candidates already in France are not eligible.

Deadline: February 2, 2007.

e-mail: Pasteurus@aol.com
www.pasteurfoundation.org

NW90578R

**SEVERAL
POSTDOCTORAL
AND PHD STUDENT
POSITIONS**

are now available in the laboratory of
Dr Lionel Berthouix at

**Université du Québec
à Trois-Rivières, Canada**

Starting date could be any time in 2007. Topics are related to host factors modulating infection by HIV-1 and other retroviruses. Salaries are guaranteed but candidates are encouraged to apply to external training grants. Prior experience in relevant fields is desirable but not required. UQTR is a French-speaking university but good English communication skills are required, and students are allowed to write their PhD thesis in English. UQTR is located in a bucolic environment, 90 min from both Montréal and Québec City. Applicants should submit cover letter, résumé, and names and contact information of 3 references, as one pdf file to be sent to:

berthouix.jobs@yahoo.com

NW90625R

Visit

www.naturejobs.com

to seriously improve
your career prospects.

naturejobs
making science work

POST DOC POSITION AVAILABLE at the Pasteur Institute (Paris), Department of Developmental Biology, Mouse Molecular Genetics Unit

A full three-year PostDoc position is available starting in January 2007 to work on the functional characterization of type 1 diabetes candidate genes in the Non-obese diabetic mouse.

The project is centered on the use of RNAi lentivirus mediated transgenesis to explore the role of the recently characterised Arnt2 gene as a causal element for the Idd6 diabetogenic loci (Hung et al. Hum.Mol.Genet. 2006). Other approaches involve transcriptional profiling, the analysis of congenic and transgenic mice, and the characterization of ex vivo models.

A knowledge of molecular biology is an absolute necessity. Some knowledge of immunology and/or mouse genetics would be an advantage.

Candidates should provide a letter of motivation, a CV, a list of publications and the names of three referees by email.

For further inquiries contact
urogner@pasteur.fr.



INSTITUT PASTEUR

W90938R

Postdoctoral Position in Cancer Research – Molecular Pathologist

A position is available in the Pathology Group, International Agency for Research on Cancer (IARC) for a candidate with a PhD or MD degree who has experience in genetic analyses of human neoplasms and a good publication record in cancer research. Current projects focus on genetic and expression profiles associated with pathogenesis of brain tumours and prostate cancer (see details at <http://www.iarc.fr>). The position is for one year with possibility of extension for one further year. Applicants should send their CV, list of publications (original, peer-reviewed papers only) and the addresses of three academic referees to:

Dr H. Ohgaki, Head, Pathology Group, IARC, 150 cours Albert Thomas, 69372 Lyon Cedex 08, France. E-mail: ohgaki@iarc.fr

W90838R



Cancer Research Scientists

Apoxis SA, an oncology-focused Swiss company based in Lausanne, has positions open in its Preclinical Oncology Group for a post-doctoral scientist and research assistant. The group's role is to investigate the potential of novel antitumour agents to be tested in clinical trials.

We are seeking applications from PhD and graduate scientists with experience in the investigation of antitumour therapies.

The successful candidates will be responsible for applying their expertise to the development of our novel, first-in-class antitumour drugs. Hands-on experience of *in vivo* tumour models is essential. The working language of the company is english.

If you would like to apply for one of these positions, please contact: **dawson@apoxis.com**

Closing date for applications: **30th November 2006.**

W90992R

Naturejobs gives an
explosive start to
your career!

naturejobs



Faculty Position in Theoretical Plasma Physics at Ecole Polytechnique Fédérale de Lausanne (EPFL)

The Center for Research in Plasma Physics (CRPP) of the EPFL is seeking applications for a tenure track assistant professor position in the field of plasma theory. We seek applicants with a strong record of scientific accomplishments, a keen interest in teaching at the undergraduate and graduate levels, and a strong potential for becoming a leader in the field. Experienced candidates seeking a higher-level position may be considered.

The successful candidate is expected to pursue independent research within the CRPP theory group, which is primarily concerned with the physics of magnetically confined plasmas for fusion. Theoretical activities in other domains related to fusion science and with interdisciplinary character are also welcome. Based on its tradition in high performance computing, the CRPP has access to state-of-the-art computer facilities.

The CRPP is part of the EPFL School of Basic Sciences and a National Competence Centre in the field of plasma physics. In addition to plasma physics teaching, its primary missions are the pursuit of scientific research in the field of controlled fusion within the framework of the EURATOM-Swiss Confederation Association and the development of expertise as well as technology transfer in the field of materials research for fusion. Activities at CRPP include a significant participation in ITER R&D and are primarily conducted in the framework of international collaborations.

Applications including a curriculum vitae, publication list, a concise statement of research and teaching interests and the names and addresses (including email) of at least five references should be submitted as a single PDF via the website <http://sb.epfl.ch/physsearch> by **January 15, 2007.**

For additional information, please contact **Professor Minh Quang Tran** (minhquang.tran@epfl.ch) or visit the following websites: <http://crppwww.epfl.ch>; <http://sb.epfl.ch/en>; <http://www.epfl.ch/Eplace.html>.

The EPFL is an equal opportunity employer.

W90742R

Postdoctoral Fellow and Studentship

We are using *C. elegans* as a model system to study cellular and molecular mechanisms underlying wiring and rewiring the brain. The *C. elegans* nervous system is composed of 302 neurons with a complete map of all axon trajectories and synaptic connections. The transparency and small size of *C. elegans* allows us to target any axon with femtosecond laser ablation, as well as visualize axonal development and axonal regeneration in live animals using time-lapse fluorescence microscopy. Femtosecond laser ablation is a new optical scalpel with exquisite precision and reproducibility. The nanometer precision of femtosecond laser ablation, as well as the million-fold shorter exposure interval, allows us to snip individual nerve fibers without collateral damages to the cell body or neighboring fibers. *C. elegans* has very powerful genetic tools to illustrate basic biological mechanisms as exemplified by *C. elegans* biologists winning 2002 Nobel Prize in programmed cell death and again this year Nobel Prize in RNAi. We thus have a unique opportunity to apply powerful genetic and optical approaches in *C. elegans* to dissect brain wiring and rewiring mechanisms. Prior experience in *C. elegans* is helpful but not essential. Applicants should have a Ph.D. degree (or an undergraduate degree if for studentship) in genetics, cell biology, molecular biology, neurobiology, biochemistry, and are familiar with molecular biology techniques. Interested candidates please send a brief cover letter describing research experience, CV, and names and contact information of two references to Dr. Chieh Chang at **chieh.chang@mcgill.ca**

Lab Website: <http://www.biology.mcgill.ca/faculty/chang/index.htm>

NW90942R



POSTDOCTORAL POSITION IGBMC - STRASBOURG (FRANCE) Dr. Michel Labouesse

To study epithelial morphogenesis in *C. elegans*.
Experience in genetics and/or cell biology requested.

For further info, see
www-igbmc.u-strasbg.fr/actualite/recrut/index.html

Please send CV to:
Dr. M. Labouesse: lmichel@igbmc.u-strasbg.fr

W90858R



BORDEAUX INSERM INSTITUTE FOR NEUROSCIENCE

The Bordeaux INSERM Research Centre in Neurosciences is a new institute funded by University of Bordeaux 2 and INSERM (The French National Institute de la Santé et de la Recherche Médicale). The centre, devoted to both research and education, is located on the campus of Bordeaux Medical School and host 6000m² of laboratory space. Research focuses on the pathophysiology of neural plasticity, ranging from neurological and behavioural pathologies to the cellular and molecular mechanisms of neural activities. Translational research in the centre aims to uncover the mechanisms of addiction, pathological pain, neurodegeneration and stress related disorders.

The centre which is already composed of 8 teams aims to become a central hub for European neurosciences. 1500m² of laboratory space are available to welcome young and talented investigators for group leader positions. Selected candidates will receive an INSERM grant and free access to the numerous core facilities of the centre including gene profiling, mice genotyping, histology and cellular imaging, animal facilities and behavioural testing.

Potential candidate please submit their curriculum to:

U862@bordeaux.inserm.fr

W90983R

**nature
cell biology**

Associate Editor

Nature Cell Biology is a prestigious international monthly journal covering all aspects of cell biology. We have an exciting opportunity available for a developmental biologist or cell biologist to join our editorial team as an Associate Editor working on all aspects of the journal.

While we are particularly interested in applicants with expertise in developmental biology and stem cells, we welcome applications from outstanding candidates in any area of molecular cell biology.

Applicants should have relevant postdoctoral experience and a strong research record. Broad scientific knowledge, excellent literary skills and a keen interest in the practice and communication of science are a prerequisite.

The successful candidate will work closely with the editor on all aspects of the editorial process, including manuscript selection, commissioning and editing of reviews and News and Views, and writing for the journal. A key aspect of the job is liaising with the scientific community and attending international conferences. The successful candidate will be dynamic and outgoing and have excellent interpersonal skills.

The Nature Cell Biology team is part of an exciting editorial and publishing environment in our central London offices which also includes the Nature and Nature Review journals. The successful applicant may be based in our established Boston or New York City offices or in London.

Applicants should submit a CV (including a brief account of their research and other relevant experience) a concise discussion in 'News and Views' format of an interesting recent scientific development, and a brief covering letter explaining current salary and quoting reference number NPG/LON/564 to: Rebecca Innes, Personnel, Macmillan Publishers, 4 Crinan Street, London, N1 9XW, or by email to londonrecruitment@macmillan.co.uk

Applications should be submitted as soon as possible.

Closing date: 23rd November 2006. nature publishing group 

90130R

NTNU – Innovation and Creativity

The Norwegian University of Science and Technology (NTNU) in Trondheim represents academic eminence in technology and the natural sciences as well as in other academic disciplines ranging from the social sciences, the arts, medicine, architecture to fine arts. Cross-disciplinary cooperation results in innovative breakthroughs and creative solutions with far-reaching social and economic impact.

The Faculty of Natural Sciences and Technology Faculty positions in Systems Biology

The positions are two of three positions in The Faculty of Natural Sciences and Technology's interdisciplinary priority area in Systems Biology. The third position will be at the Department of Biotechnology.

Associate Professor in Biology (Systems Biology) – ref.no NT-40/06

Enquiries about the position may be made to the Head of Department, Professor Eivind Røskoft, tlf. + 47 73 59 60 73, e-mail: roskoft@bio.ntnu.no or Deputy Head of Department, Berit Johansen, tel +47 73596107, e-mail: bejo@bio.ntnu.no Further information on the position is available at: <http://www.bio.ntnu.no>

Professor/Associate Professor in Chemical Engineering (Systems Biology) – ref.no 42/06

Enquiries about the position may be made to the Head of Department, Professor Sigurd Skogestad, tlf. + 47 73 59 41 54, e-mail: sigurd.skogestad@chemeng.ntnu.no Further information on the position is available at: <http://www.chemeng.ntnu.no>

Applications should be sent to the Norwegian University of Science and Technology, Faculty of Natural Sciences and Technology, N-7491 Trondheim, Please quote NT-40/06 (Biology) or NT 42/06 (Chemical Engineering) in all correspondence.
Closing date 2006-12-01.

Complete announcement text on NTNU's website:
http://nettopp.ntnu.no/?kat=N_JOB

 **NTNU**
Innovation and Creativity

W90614R

Making the sale

Give 'em the easy choice.

Fredric Heeren

It was that Presbyterian biologist who almost blew the sale. And not only a sale to the Presbyterians, but to the Seventh-Day Adventists, too. Those two denominations alone could deliver 18 million customers, guaranteed. It was a good thing I came along with the sales staff for this one. Half-way through my product-design talk, the biologist interrupted, as though suddenly possessed, and began telling the others they were making a big mistake.

I brought the denominational leaders back to the purpose of the meeting when I showed them what our Personal Advice Device could do.

Of course James had done his usual great job of setting the hook. Nobody could argue with his facts. People have come to trust their PADs implicitly. More than their spouses. More than their pastors. What human being can compete with this advice from a mind programmed to think according to our individual tastes, but immeasurably smarter, continuously updated from a world of information according to our present needs?

So here's how I explained why our product is the solution to their evolution issue — and if you pay close attention, you'll know how to handle the Mennonites next week.

"If you want to look at the future of your denomination," I said, "you only have to look at how the Southern Baptists were transformed from criticizing evolution in the schools to promoting it. This was bound to happen, with all the preachers telling their people there's no evidence that primitive hominids ever existed.

"That might have been fine in the day when, to check out that claim, a person would have had to sit down at a PC and hunt around for hours — who has time for that? But now for those who have PADs, questions like that get answered with just a few thoughts and a quick menu selection on their retinas. Presto — they're looking at a sequence of hominid photos with increasing cranial capacities over time. And seeing is believing, even if you've been home-schooled and never heard of evolution except as a naughty word. The result: anti-evolution leaders were voted out by an informed electorate.

"If our projections hold, 80% of American teens and adults are going to own a

PAD within three years. There's no choice about that, but what you can still choose, courtesy of our latest design and your specs, is whether your PAD will be your enemy or your friend.

"We've designed a *friendly* PAD. It's not just a matter of helping people make good decisions by showing them their best options — it actually does the rationalizing for them, too. This is as close to real human thinking as any PAD has ever come."

Now at this point, I thought we were still going according to plan. The Presbyteri-

anologist dismissed it — and everyone knows a PAD's logic is flawless. It also showed an alternative, God-honouring explanation for the putative evidence. It suggested that several of the questions were based on false preconceptions. And always, it directed people to look at their denominational virtwebs, where they would get the true picture. Of course, we restricted access to the problem virtwebs.

"So," I concluded, "it really operates exactly as the human mind does. It chooses what it wants to believe."

That's when the biologist piped up and said something about our spoiling the whole point of genuine choice: young people shouldn't be made to choose between faith and science.

"Of course not," his leaders said. "Not as long as you're talking about *Christian* science."

"We're seeing the Christian Science people next week," said James. And from there things got very confusing.

The Presbyterian biologist got into an in-house discussion with his leaders about whether the particular means God used to create our bodies should be the issue.

"It certainly *is* the issue," the Presbyterian president said.

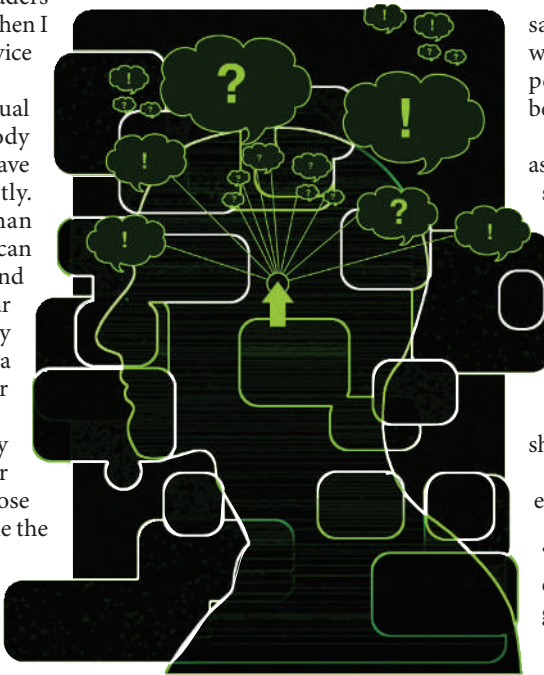
"It shouldn't be," said the biologist, "because the more genetics or history our bodies share with other animals, the greater the wonder at what we humans uniquely experience: morality, humour, literature, science, faith."

"Dr Adams," said the president, "you were invited here to ask your questions and you've done that."

"But I have more," said Adams. "Do we really want to credit all these attributes merely to something special about our bodies, as if the material world is all there is? Isn't that what we're doing when we emphasize the special way our bodies were made? Isn't that what we're doing when we pit God against evolution?"

Some of the denominational leaders began to show a hint of worry. Fortunately, his remarks had triggered their PADs, which were already fast at work drowning him out with a flood of dazzling answers, backed up by well-documented evidence and memorable sound bites.

Fredric Heeren is a researcher and science writer in Kansas. By the time he finishes his book on human evolution, he fears we may have evolved into something else.



JACEY

ans had brought their biologist to give me the chance to demonstrate how our PAD works. I let him fire away with his toughest questions.

"How do you explain the stark differences between animals across impassable barriers, as between Australia and other lands? Why do islands always have organisms most like the nearest mainland? Why do we find clear evidence of common ancestry when we compare the DNA of an evolutionary sequence of animals, so that we can trace how genetic insertions, including retroviruses, appear and then accumulate in later forms of life?"

The Adventists were set up for PADs and got answers in their heads. I showed the others on the large virtweb display what our PAD does compared with the others. As soon as any question was seen to contradict a specified tenet, our PAD

Abstractions



LEAD AUTHOR

Roger Bilham is on a mission to improve earthquake forecasts in the Himalayan region. The geophysicist and his student Nicole Feldl, from the University of Colorado,

Boulder, braved high altitudes and Maoist guerillas to gather GPS points in the Himalaya. Using these data and a subsurface model of the northern Indian region helped clarify the seismic significance of the squishy Tibetan plateau that has confounded modelling efforts up to now. On page 165, the model represents how time lag between earthquakes, the length of ruptures and their magnitude are linked.

How soon could the Himalaya experience a massive earthquake?

Our calculations show that two-thirds of the Himalaya are ready today for an earthquake of magnitude 8 or greater. Some parts of the Himalaya are apparently long overdue.

Describe your model.

By assuming that most of the Tibetan plateau is elastic, or able to return to its original shape following temporary stress, we calculated how earthquake magnitude grows with rupture length. Using data from past earthquakes, we calibrated our results by adjusting the time interval in our models that's necessary to obtain the observed geologic movement. Unfortunately, it's not useful for accurately predicting earthquakes.

How does this work change assumptions about plate activity in the region?

It brings a new perspective, with a couple of unexpected consequences. For example, if you have a devastating magnitude 7.6 earthquake (like in Kashmir, 2005), we find that it doesn't drain the system's energy, and the same area can have another, bigger earthquake soon afterwards. This contradicts seismic gap theory, which says if you've recently had a big earthquake you shouldn't expect another for a long time.

If pressure is released by smaller earthquakes, how can conditions for a megaquake still exist?

Short ruptures are unable to tap deep into the Tibetan plateau's cumulative reservoir of strain energy. It requires megaquakes to do this. We've had four fairly substantial earthquakes in the past 200 years, but they haven't been big enough to release much of this cumulative strain.

How serious is the threat of earthquake in the region?

Quite frankly, the countries bordering the Himalaya are in terrible trouble. We estimate that one million people could die in a single event unless city structures are made earthquake-resistant. ■

MAKING THE PAPER

Ya Ha

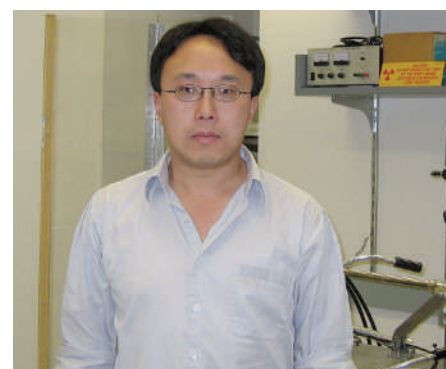
Understanding how proteases cut inside the cell membrane.

Scientists have long puzzled over the workings of membrane-spanning proteases, the molecular scalpels that cleave other membrane proteins into smaller segments. Their interest is due, in part, to the fact that this group of enzymes includes γ -secretase, the protease responsible for producing the amyloid β -peptide that forms damaging plaques in the brains of Alzheimer's sufferers. Understanding how γ -secretase works may lead to ways to stop its function and hinder the disease. On page 179 of this issue, Ya Ha of Yale University School of Medicine, New Haven, determines the crystal structure of another intramembrane protease from *Escherichia coli*. Although unrelated to γ -secretase in amino-acid sequence, the protease he describes is likely to act in the same way.

Proteases cleave proteins into smaller segments using water molecules to break amino-acid bonds. Thus, they typically operate in a watery environment. But the proteases embedded in the cell's membrane are an exception. They do the cutting inside the cell membrane.

Ha heard about this group of enzymes as a postdoc at Harvard University five years ago while searching for a project for his own lab. The intramembrane proteases seemed to fit the bill. "This area was ideal for a structural biologist like myself because there is no other way to figure out how these proteases work," he explains. "You need direct visualization to do so."

After setting up his lab at Yale in 2001, he set out, with the help of Yongcheng Wang and Yingjiu Zhang, to determine the crystal structures of several intramembrane proteases, including γ -secretase. He followed standard laboratory practice, engineering the genes encoding these proteases so that they were



expressed in bacteria. He purified each protein to make crystals from it, and then bombarded the crystals with X-rays to reveal the relative positions of the protein's atoms.

"We were stuck at almost every step of the process," laughs Ha. "It was a risky project." Most of the proteins he had started with had to be discarded at various stages of the purification process because they did not yield either sufficient amounts or sufficiently pure or stable protein. But he persevered. "Structural biologists have a number of tricks to use," he says. In the end, he was able to crystallize a bacterial rhomboid protein called GlpG.

With crystals in hand, he quickly obtained the protein's structure. Ha discovered that the amino acids involved in cleaving target proteins reside in a cavity smack in the middle of the protein, along with several water molecules. Although the membrane keeps water out, inside the enzyme there's a little watery pool where the proteins are cut. In the absence of a protein, the opening to this cavity is blocked by a string of amino acids shaped into a loop, which may act as a molecular gate.

Ha plans to use his experience to unravel the structure of other intramembrane proteases, including γ -secretase. "Now we know how difficult the project is, but that it can be done," he says, while acknowledging that the process may not be any easier the second time around. "I have been in the field long enough to know that experience does not always translate into making things go faster." ■

KEY COLLABORATION

Chimpanzees are recognized as the primary primate reservoir for simian immunodeficiency viruses (SIV) — the most closely related virus to HIV-1, which causes AIDS in humans. Earlier this year, the origins of both a pandemic and non-pandemic form of HIV-1 were traced to distinct chimpanzee groups in southern Cameroon. The same research team found that wild gorillas in western Cameroon were infected with a variant of SIV that is more

closely related to a strain of HIV-1 (see page 164).

The ten-year collaboration of researchers — from France, the United States, the United Kingdom and Cameroon — collected fecal samples from chimpanzees and gorillas in the remote regions of Cameroon. "It is key to have a local team involved in such research efforts," says Martine Peeters, a virologist at the University of Montpellier, France.

The samples were sent to

labs in France and Alabama equipped to extract fecal RNA.

Peeters stresses that the gorilla findings don't negate previous chimpanzee work. "Chimpanzees could have transmitted the newly found virus to gorillas and humans independently, or they could have been transmitted first to gorillas, who transmitted it to humans," she says.

The team will soon perform analyses at a molecular-biology lab to be built in Cameroon. ■